

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 09:31.07 ; Search time 3864 Seconds

(without alignments)
3856.105 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAVQWTRATEPRYP.....WDEPRKLLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model
-Q=/cgn2_1/USFTQ.epool_p/US09522733/runat_02112005_173640_20785/app_query.faetca_1.2695
-DB=N.Geneseq.16Dec04 -QFMT=faetcap -SUFTX=p2n.rng -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNIT=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NCRM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09522733@cgn_1.1.1431@runat_02112005_173640_20785 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.16Dec04:.*
1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13215	100.0	7554	12	ADJ92815 Human co-
2	13215	100.0	8561	12	ADG86298 Human SMR
3	13215	100.0	8561	12	ADN04304 Anticpsoct
4	13178.5	99.7	8564	3	AACT4783 Human ORF
5	13145	99.5	8667	11	ACN44283 Human mRN

6	13119.5	99.3	8533	13	ADG84524 Human tum
7	13119.5	99.3	8533	13	ACN39603 Tumour-as
8	12978	98.2	7521	8	ACA62250 Human nuc
9	12978	98.2	7524	10	ADL13812 Osteoarth
10	12978	98.2	8686	8	ACA62249 CDNA enco
11	12978	98.2	8686	10	ADL13811 Osteoarth
12	12978	98.2	8686	12	ADG86290 Human SMR
13	12978	98.2	8686	12	ADQ18920 Human sof
14	12922	97.8	9053	12	ADL12577 Human sof
15	12845.5	97.2	9079	12	ADQ23294 Human sof
16	10832.5	82.0	7386	8	ACA62452 Mouse nuc
17	10832.5	82.0	8544	8	ACA62451 CDNA enco
18	9691.5	73.3	7534	11	ACN44281 Mouse mRN
19	7885	59.7	5989	6	ABK84305 Human CDN
20	4612	34.9	2930	12	ADG83959 Human brc
21	4612	34.9	2930	12	ADG86301 Human brc
22	4187	31.7	7780	3	AAA60630 HNRCR nuc
23	4165.5	31.5	7914	12	ADG83959 Human tum
24	4165.5	31.5	7914	13	ADG87683 Human tum
25	4147.5	31.4	7940	8	ABZ34833 Coding se
26	4147.5	31.4	7940	13	ADG89791 Antagonis
27	4043	30.6	7900	3	AAA60629 Human HNR
28	3649	27.6	91141	11	ACN44280 Mouse gen
29	2951	22.3	23380	11	ACN44282 Human gen
30	2184	16.5	2745	12	ADJ92816 Human co-
31	2133.5	16.1	220756	12	ADG86300 Human SMR
32	1871.5	14.2	2336	10	ADK31306 Human dia
33	1391	10.5	2914	13	ADR07585 Full leng
34	1113	8.4	650	8	ACA57401 Human adi
35	1099	8.3	10410	13	ADG56633 Drosophi1
36	1083	8.2	10910	4	ABL03131 Drosophi1
37	1076	8.1	872	3	AAA02670 Human col
38	1073.5	8.1	10900	13	ADG89593 Antagonis
39	977.5	7.4	2519	12	ADP48942 Human-yea
40	960	7.3	555	8	ACA57524 Human adi
41	880	6.7	718	10	ADG76358 Human BSK
42	845	6.4	710	4	AAI197539 Human neu
43	774.5	5.9	673	8	ACA57523 Human adi
44	746.5	5.6	520	6	ABK44934 CDNA enco
45	726.5	5.5	33529	5	AAI17367 DNA seque

ALIGNMENTS

RESULT 1	ADJ92815	ADJ92815 standard; DNA; 7554 BP.
XX	ADJ92815;	
AC	ADJ92815;	
XX		
DT	06-MAY-2004	(first entry)
XX		
DE	Human co-repressor SMRT DNA.	
XX		
KW	Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;	
KM	atherosclerosis; human; co-repressor; gene; de.	
XX		
OS	Homo sapiens.	
XX		
PN	US2003228607-A1.	
XX		
PD	11-DEC-2003.	
XX		
PF	14-APR-2003; 2003US-00414692.	
XX		
PR	15-APR-2002; 2002US-0372650P.	
PA	(WAGN/) WAGNER B L.	
PA	(SCHU/) SCHULMAN I G.	
XX		
PI	Wagner BL, Schulman IG;	
XX		
DR	WPI; 2004-167207/16.	

XX Identifying compounds that bind to nuclear receptor and exhibit cell type
PT specific actions, and useful for treating hyperlipidemia, obesity and
PT diabetes.
XX
XX Claim 5; SEQ ID NO 14; 99pp; English.
XX
XX The invention relates to screening methods for identifying compounds that
CC bind to nuclear receptor and exhibit cell type specific actions. The
CC invention relates to modulators having an improved therapeutic profile.
CC The method is useful for identifying compounds that bind to a nuclear
CC receptor and exhibit cell type specific actions. It is also useful for
CC identifying modulators of nuclear receptors that are useful in treating
CC diseases e.g., diabetes, hyperlipidaemia, obesity, atherosclerosis, etc.
CC The present sequence is human co-repressor DNA used to illustrate the
CC method of the invention.

SO Sequence 7554 BP; 1646 A; 2732 C; 2228 G; 948 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	7554
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-09-522-753-5 (1-2517) x ADJ92815 (1-7554)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
DB 1 AAGTCGGGCTCCACACAGCTTGTCGACAGACGTGAGGGCCACTGAGCCCGCTACCGG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 40
DB 61 CCCACAGGCTTTCCTACCAAGTCAGATCGCCCGACGACGACGACGCGGGCTCTCG 120
QY 41 GlnTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACACACACACTCCGCGACTGCTGCCACTGCCCTGCGCGGCTTCATCATCCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyIleAsnGluArgSerGln 80
DB 181 CCCACAGGCGGAGGGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGAAATGAACGCTCCAG 240
QY 81 GlnLeuHisLeuArgProGlySerHisSerTyrLeuProGlyIleGlnIleGlnMet 100
DB 241 GAGCTCCACTCGGCGGACAGTCCCACTACCTGCGGAGCTGGGGAAGTCAGAGATG 300
QY 101 GlnPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAAGAGCTGCTGACCCCTGCTGCGAGCG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAsnLeuThrLysAspArgSer 140
DB 361 TCACCCCTCTGCGCCAGCGGCGCAGCTCGGGATTCGAAGACTTCACCAAGGACCGTAGC 420
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGlyLeu 160
DB 421 CTGACGGGGAAGCTGGAACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerLysGlnGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGGTCCGCCACGCGCTGTCCAAAGAGAGAGACTGATCCAAACATGAGCCGCTGAGAC 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 541 CCAAGATATCAACATGGTAGAGAGAGATCTTAACTGAAGAAAGAGAGAGAGAGAGCTG 600
QY 201 GlnGlnGlnAlaAlaLysProProGluProGluLysProValSerProProIleGln 220
DB 601 GAGGAGAGAGGCTCCAAAGCCGCGGAGCTGAGAAAGCCGCTGTCAACCGCCCATCGAG 660

QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysAlaGlnAla 240
DB 661 TCGAAGCACCGCAGCGCTGTGTGACATCTACGACGAGAACCGGAAGAGCGTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnValGlnLeuProLeuThrAsnGlnPro 260
DB 721 GACATCGGATTCGTGAAAGGCTTGAGGCCCCAGGTGAGACTGCGGTGTACAAACGAGCC 780
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
DB 781 TCGACACCCGGCAGATCATGAGAACATCAAAATTAACAGGCGATGCGGAAGAGCTA 840
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnThrLysGlnLysPheCysGln 300
DB 841 ATCTGTACTTCAGAGGAGGATCACTGCGAAACAATGAGAGAGAGATTCGTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysLysValGlnLysIleGlnAsnAsnPro 320
DB 901 CCTATGACCAAGCTCATGAGAGCTTGGAAGAAAAAGTGAAGCGCATCCAAAACACCCG 960
QY 321 ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle 340
DB 961 CCGCGGCGGCGCAAGAGAGCAAGGTGCGGAGTACTACGAAAGCAGTTCCCTGAGATC 1020
QY 341 ArgLysGlnArgLysGlnGlnGlnLysArgMetGlnSerArgValGlnIleArgLysSerGly 360
DB 1021 CCGAAGCAGCGGAGCTGAGAGCGCATGACAGAGAGGCTGGCCAGCGCGGCGAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleLeuAspGlyLeuSer 380
DB 1081 CTGTTCATGTCCGCGCCCGCAGCAGCAGCATGAGAGTGTACAGATCATTCATGCGCTCTCA 1140
QY 381 GlnGlnGlnAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1141 GAGCAGAGAACCTGGAGAAAGAGATGCGCAGCTGCGCGGTATCCCGCCCATGTGTAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1201 GACGCTGACACACAGCGCATCAAGTTCACTCAATGAAGGGCTTATGGCCGACCCCATG 1260
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGlnGlnLysGlnThrPhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCTATGAACATGTGAAGTACAGACAGAAAGAGACTTC 1320
QY 441 ArgGlnLysPheMetGlnHisSerProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CCGAGAAAGTTCAATGACAGCATCCCAAGAACTTTGGCTGATCGCATCTCCGAGAGG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnAsnTyrLys 480
DB 1381 AAGACAGTGTGAGTGTGCTCTCTATTACTGACTGACCTGAAGAAATAGAACTTAAAG 1440
QY 481 SerLeuValArgAspSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
DB 1441 AGCTGTGTAGAGAGAGCTGTATGCGGCCCGCGCAGAGCAGCAGCAGCAGCAGCAG 1500
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
QY 521 AspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 540
DB 1561 GATGAGAAAGAGAAAGAAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 541 AspLysGlnAspLeuLeuLysGlnLysThrAspAspThrSerGlyGlnLysAsnAspGln 560
DB 1621 GACAAAGAAAGACTCTCTCAAGAGAAAGACAGACGACACTTACGGGAGAGACAGAGAG 1680
QY 561 LysGlnAlaValAlaSerLysGlnArgLysThrAlaAsnSerGlnLysArgArgLysGly 580
DB 1681 AAGAGAGCTGTGGCTCCAAAGCGCGCAAAACTGCCAAACAGCAGGAGAGAGAGAG 1740
QY 581 ArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnLysAlaIleThrProGlnGln 600

Db	1741	CGCATCACCCGCTCATAGCTATATAGGCAACAGCGAGAGGCCATCACCCCAGCAG	1800
Qy	601	SerialgluLeuAlaSerMetGluLeuLeuGluSerSerArgTrpThrGluGluGluMet	620
Db	1801	AGCGCGAGCTGGCTCTCCATGGAGCTGAATGAAGATTCTCGCTGGACAGAGAAAGATG	1860
Qy	621	GluThrAlaLysLysGlyLeuLeuGluLysGlyArgMetTrpSerAlaLeuAlaArgMet	640
Db	1861	GAAGACAGCCAGAAAGAGTCTCTCGGAACACGGCCCAACTGGTGGCCATCCCGCGATG	1920
Qy	641	ValGlySerLysThrValSerGlnCysLysAsnPhenylrPhenylrLysLysArgGln	660
Db	1921	GTGGGCTCCAAAGACTGTGTGCGAGTGAAGACTTCTACTCACTTACAGAAAGAGCAG	1980
Qy	661	AsnLeuAspGluLysLeuGlnGlnLysLysLeuLysMetGluLysGluLysArgAlaArg	680
Db	1981	AACTCGATGAGATTCTTGACAGACAGCAAGCTGAAGATGAGAGAGAGAAAGAACCGCGG	2040
Qy	681	ArgLysLysLysLysAlaProAlaAlaAspGluGluAlaAlaPheProProValVal	700
Db	2041	AGGAAGAAAGAAAGAACCGCGCGCGCGAGCGAGGCTGCAATCCCGCCGCTGGTG	2100
Qy	701	GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu	720
Db	2101	GAGGATGAGGATGAGAGCGCTCGCGCGTGAAGCGGAAATGAAGAGAAATGCTGAGAGAG	2160
Qy	721	AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla	740
Db	2161	GCTGAAGCTTTCATAGCTCTGGAGAAATAGGTGCCAAGGGGAAATGCAGTGGCCACCC	2220
Qy	741	ThrValAsnAsnSerSerAspThrGluSerLysProSerProHisLeuThrGluAlaLys	760
Db	2221	ACTGTCAACAAACAGTCTCAGACACCGAGAGCATCCCTCTCTCTCACTGAGGCGCCAG	2280
Qy	761	AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyValAspGlyProProPro	780
Db	2281	GACACAGGGCGAAGAGGGCGCCAGCGCCAGCCACTGGGGCGCGAGCGGCCACCCCA	2340
Qy	781	GlyProProThrProProArgArgThrSerArgAlaProLysGluProThrProAlaSer	800
Db	2341	GGCCACCCACCCCAACAGGAGACATCCGGGCGCCCATTTGAGCCCAACCCCGGCTCT	2400
Qy	801	GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro	820
Db	2401	GAAGCCACCGGAGCCCTTACGCCGCCACAGCAACCCCATGCTCTGCACTCTCTCT	2460
Qy	821	ValValProLysGlyGluLysGlyGluGluThrAlaAlaAlaProProValGluGluGly	840
Db	2461	GTGGTCCCCAAGAGAGAAAGAGAGGAGACCGGACCGGCCCCCACTGAGAGAGGG	2520
Qy	841	GluGluGluLysProProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu	860
Db	2521	GAGGAGCGAAGAGCCCGCGCGCTGAGAGAGCTGGGACGTGACACAGGAGAAAGCCGAGAG	2580
Qy	861	ProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAla	880
Db	2581	CCCGTCMAAGACGAGTGCACGGAGGAGCCAGAGAGGGGCGCGCAAGGGCAAGAGCCG	2640
Qy	881	GluAlaAlaGluAlaThrAlaGluGlyValaLeuLysAlaGluLysLysGlyGlySer	900
Db	2641	GAGGCGCGTGAGGCGCACGGCGGCGGCGCTCAAGGACAGAAAGAAAGAGAGGGGAGGC	2700
Qy	901	GlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThr	920
Db	2701	GGCAGGGGCGACACTGCGCAAGAGCTCGGGCGCCCGCCAGGACAGGACTTCCAGTGTACC	2760
Qy	921	CysSerAlaAspGlyValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro	940
Db	2761	TGCAGTGCAGACGAGGTGATGAGGCCAGAGGGCGGCGCAAGAACCGGCTCTGTCCCA	2820
Qy	941	ArgProSerLeuLeuThrProThrLysAspProArgAlaAsnAlaSerProGlnLysPro	960
Db	2821	AGGCCACCGCTCTTCACCCCGCATGGCGGACCCCGGGCCAAATGCCCTCACCCCAAGCCA	2880
Qy	961	LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaLeuProProLysGlnValThrLys	980
Db	2881	CTGGACCTGAAGACGCTGAAGACAGAGCGGCTGCATCCCCCATTCAGAGTCAACCAA	2940
Qy	981	ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro	1000
Db	2941	GTCCATGAGCCCCCGGAGAGAGCGAGCTCCCAACCAAGCCAGTCCCCCGAGCCCAACG	3000
Qy	1001	ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg	1020
Db	3001	CAACCGCAAACTTGACAGCCGGAAGAGGACCCCTTCAGACAGCTTGAGAGAGCCCCCG	3060
Qy	1021	GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln	1040
Db	3061	GGCAAGAGCAGAGACCCCGGACCCCGCGCGCAAGAGAGGCTTGAGAGCGGAGGCCAG	3120
Qy	1041	LysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArg	1060
Db	3121	AAAGCTGCTGGAGACCCCTTGTGTGAGCTTCGGGCTGCGCTTCCCGTCCCCCGCT	3180
Qy	1061	GluValLysLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProPro	1080
Db	3181	GAGGTGATCAAGGCTCTCCCGCAAGCCCGGACCCCTCAGGCTTCTCTTACGTCCACT	3240
Qy	1081	GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro	1100
Db	3241	GGTACACCACTGGCCCTTGAGCTCATGACATGCCCCGGCGGCTCTGGCGGCCACCC	3300
Qy	1101	ThrLysSerAsnProProProLeuLysSerSerAlaLysHisProSerValLeuGluArg	1120
Db	3301	ACCATCTTCAACCCGCTCCCTCATCTCTGTCGCAAGACCCCAAGCTCTTCAAGAG	3360
Qy	1121	GlnLysGlyAlaLysSerGlnLysMetSerValGlnLeuHisValProTySerGluHis	1140
Db	3361	CAATATAGTGCATCTCCCAAGAAATGTGGTCCAGTCCACGTCCTGATCAGACAT	3420
Qy	1141	AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	3421	GCCAGGCCCCGGTGGGCTCTGTACCATGGGGCTGGCCCTTGCCATGAGACCCCAAAAG	3480
Qy	1161	LeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro	1180
Db	3481	CTGGCACCTTACGCGGAGTGAAGAGACAGAGAGCTGCCAAGGGGCGAGGCTGGGCA	3540
Qy	1181	ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu	1200
Db	3541	CCGAGAGAGCTTGGGGGTGCCACAGCCCAAGAGCGTCTGTGAGAGAGGACAGCTCTG	3600
Qy	1201	GlySerValProGlyGlySerLysLeuThrGlyLysLeuSerThrArgValProSerAsp	1220
Db	3601	GGCTCACTTCCGGCGGAAACATCACCAAGAGCATTCCAGCACACGGGTGCTCGGAC	3660
Qy	1221	SerialLeuThrTyraArgLysSerLysThrHisGlyThrProAlaAspValLeuTyLys	1240
Db	3661	AGCGCATTCATACCGCGGCTCATACCAACGAGCAGCGCACTGACGTCTGTACAG	3720

QY 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyValArgAlaIleProProGluIuaArgHis 1340
DB 3961 GCCATCTCTCAAGCCAGCATCGAAGGTCTCATGTGGCGGTGCATCCCGCGGAGCCAGAC 4020
QY 1341 SerProHisHisIleuIuysGluGluGlnHisIleIaArgGlySerIleThrGlnGlyIlePro 1360
DB 4021 AGCCCCACCACTCAAGAGCAGCACACATCCGGGGGTCCATGCACACAGGGGAATCCCT 4080
QY 1361 ArgSerTyValGluIuaGlnGluAspTyLeuIuaArgAlaGluAlaIysIleuIuysArg 1380
DB 4081 CGGTCTTAGTGAGGACACAGAGACTACCTCGCTCGGAGGCCAAAGCTCTCAAGCCG 4140
QY 1381 GluGlyThrProProProProProSerArgAspIleuThrGluAlaTyIleThrGln 1400
DB 4141 GAGGGACCGCTCCGCCCAACCGCCCTCACGGGACTGCACGAGGCTTACAAAGACGAC 4200
QY 1401 AlaIeuglyProIeuIuysIleuIuysProAlaHisGluGlyLeuValAlaThrValIysGlu 1420
DB 4201 GCCCTGGGGCCCCGTGAAGCTGAAGCCGCCCATGAAGGGCTGGTGGCCACGGTGAAGAG 4260
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluIleuIuysThrProGluIePro 1440
DB 4261 GCGGGCGCTCCATCATGAGATCCGCGGAGAGACTGGGACACGCCCGAGCTGCC 4320
QY 1441 LeuAlaProArgProIeuIuysGluGlySerIleThrGlnGlyThrProIeuIuysTyAsp 1460
DB 4321 CTGGCCCCCGCGCGCTCAAGAGAGGCTCATCACAGGAGCACCCGCTCAAGTACGAC 4380
QY 1461 ThrGlyAlaSerThrThrGlySerIleIysHisAspValArgSerIleuIleGlySerPro 1480
DB 4381 ACCGGGGCGTCCACCACTGGCTCAAAAGACAGACGATAGCTCCCTCATCGGACACCCC 4440
QY 1481 GlyArgThrPheProProValHisProIeuAspValMetAlaAspAlaArgAlaIeuglu 1500
DB 4441 GCGCGGACGTTCCACCGCTGCACCGCGTGGATGTATGACCCACCGCCGGGACATGGAA 4500
QY 1501 ArgAlaCysTyArgIleuIuysIleuIuysSerArgProGlyThrAlaSerSerSergIyGly 1520
DB 4501 CGGCTGTGCTACGAGAGAGCTGAAGAGCCGCGCACGAGGACCGCCAGAGACTGGGGGG 4560
QY 1521 SerIleAlaArgGlyAlaProValIleValProGluIeuglyIysProArgGlnSerPro 1540
DB 4561 TCATATGCGCGCGCGCCCGCTCATTTGCTCGAGCTGGTAAAGCCGCGGACGACCCC 4620
QY 1541 LeuThrTyGluAspHisGlyAlaProPheAlaGlyHisIleuProArgGlySerProVal 1560
DB 4621 CTGACCTTAGAGACACAGGGGACCCCTTTGCCGGCACCTCCACGAGGTTGCCCGG 4680
QY 1561 ThrMetArgGluProThrProArgLeuGlnGlySerIeuSerSerSergIyAlaSer 1580
DB 4681 ACCATCGGGAGCCACGCGCGCTGCAGAGGAGGACCTTTGCTCGACAAAGCATCC 4740
QY 1581 GlnAspArgIysIleuThrSerThrProArgGluIleAlaIysSerProHisSerThrVal 1600
DB 4741 CAGGACCGAAGCTGACGTGCAGCTCTGTGATGATGCCCAAGTCCCGGCACACACCGTG 4800
QY 1601 ProGluHisHisProHisProIleSerProTyGluHisIleuIuysArgGlyValIleSergly 1620
DB 4801 CCGGACACACACACACACCCCATCTGCCTTAGAGACTGCTGGGGCGGAGGGG 4860
QY 1621 ValAspLeuTyArgSerHisIleProIeuaAlaPheAspProThrSerIleProArgGly 1640
DB 4861 GTGGACTGTATGAGGACACATCCCGCTGGCTTGACCCCACTCCATACCCCGGG 4920
QY 1641 IleProIeuAspAlaAlaAlaAlaTyTyTyLeuProArgHisIleuAlaProAspProThr 1660
DB 4921 ATCCCTCTGACACACACCGCTGCTACTACTGCCGACACCTGGCCCCCAACCCACAC 4980
QY 1661 TyTyProHisIleuTyTyProProTyTyIleuIleArgGlyTyTyProAspThrAlaAlaIeuglu 1680
DB 4981 TACCCGACCTGTACCAACCTTACTTCATCCGGGCTACCCGACACAGGGGCGCTGAG 5040

QY 1681 AsnArgGlnThrIleIleIleAsnAspTyTyIleThrSergGlnGluMetHisHisAsnThrAla 1700
DB 5041 AACCGGACACCATCATCTATCATATCATCATCTGCACAGATGTGACCAACAGCGGCC 5100
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSergLeu 1720
DB 5101 ACCGCGATGGCCAGGAGACTGATATGCTGAGGGGCTCTCGCCCGCGAGTCTTGCTG 5160
QY 1721 AlaIeuAsnTyAlaAlaGlyProArgGlyIleIleAspIeuSerGlnValProHisIleu 1740
DB 5161 GCACCTCACTACGCTCGGGTCCCGGAGCATCATGACTGTGCCAATGTGCACACCTG 5220
QY 1741 ProValIeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
DB 5221 CCGTGCTGTGGCCCCCGACACAGGACCCACAGCCACCATGAGACGCTTGCTGCTAC 5280
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProIeuSerProGly 1780
DB 5281 CTCCCAACCGCGCCCGACGCTTCAGACAGCCGCCACAGACTCCCACTTCCCAAGA 5340
QY 1781 GlyProThrHisIleuThrIysAspProThrThrThrSerSerSergIuArgIuaArgAspArg 1800
DB 5341 GGTCCAAACACTGTACAAACCAACCAACCAACCAAGTCTCGTCCGAGGGAGCGAGACCG 5400
QY 1801 AspArgGluArgAspArgAspArgGluArgGluIysSerIleIleuThrSerThrThrThr 1820
DB 5401 GATCGAGACCGGAGCCGGGATCGGAGCGGAAATCATCTCATCGTCCACACAGACG 5460
QY 1821 ValGluHisAlaProIleThrArgProGlyThrGluGlnSerSergIySerSergIySer 1840
DB 5461 GTGGAGACGACCACTCTGAGACTGTGTACAGACAGACAGACAGCGGACGCGGACG 5520
QY 1841 SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5521 AGCGGGGGGGGGGGGAGCAGACAGCGCGCGCTCCCATCCCATGCCACACAGCAC 5580
QY 1861 SerProIleSerProArgThrGlnAspAlaIeuglnGlnArgProSerValIeuHisAsn 1880
DB 5581 TCGCCCATCTCCCTCGAGCCAGATGCTTCACAGACAGACCCCATGCTGTTCACAA 5640
QY 1881 ThrGlyMetIysGlyIleIleThrAlaValGluProSerIysProThrValIeuAspSer 1900
DB 5641 ACGGCGATAGGGATATCATACCGCTGTGAGGCCAGAGCCACCGGTCTGAGGTCC 5700
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5701 ACTTCACCTCTCTACACCGCTTGCCGACGTGCACATTCACACTGCGACCCACCTGCCA 5760
QY 1921 LeuGlyGlyThrIeuAspGlyValTyTyProThrLeuMetGluProValIleuIeuProIys 1940
DB 5761 CTGGGCGGACCTTCATGGGGGTCTTACCTTACCTCATGAGAGCCGCTCTTGCTGCGCAAG 5820
QY 1941 GluAlaProArgValAlaAspProGluIuaArgProArgAlaAspThrGlyHisAlaPheIeu 1960
DB 5821 GAGGCCCCCGGGGTGCGCGGCGAGGCGGCCCGAGACACGCGCATGCTTCTTC 5880
QY 1961 AlaIysProProAlaArgSergIyLeuGluProAlaSerSerProSerIysSergIu 1980
DB 5881 GCCAAGCCCCCAGCCCGCTCGGGGTGAGCCGCGCTCTCCCGCAGCAAGGCTCGAG 5940
QY 1981 ProArgProIeuValProProValSergIyHisAlaThrIleAlaArgThrProAlaIys 2000
DB 5941 CCGCGGCCCCGTAGTGTCTGTCTGTGCGACAGCCACATGCGCCGACCCCTGGGAAG 6000
QY 2001 AsnIeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
DB 6001 AACCTCGACCTTACACACCGACCGGACCGGCGGCGCACCTGCGGCTGAGAC 6060
QY 2021 ProHisArgGluIysThrGlnSerIysProPheSerIleGlnGluIeuglyIeuAspArgSer 2040
DB 6061 CCGGACCGGAAAGACTCAAGATAACCTTTTTCATGTCAAGAACTGGAATCTCGTTCT 6120
QY 2041 LeuGlyTyTyHisGlySerSerTySerProGluGlyValGluProValSerProValSer 2060


```

Db      6121 CTGGGTTACCAACGAGCAGCTACAGCCCGAAGGGGTGAGCCCGCTGACCCCTGTAGC 6180
Qy      2061 SerProSerLeuThrH1aApLyAG1yLeuProLyH1aLeuG1uG1uLeuAAspLySer 2080
Db      6181 TCACCCAGCTGACCCACGACAAAGGGCTCCCAAGACCTCGAAGAGCTGACAAAGGC 6240
Qy      2081 H1aLeuG1uG1uLeuAAspLyAG1uProLyG1uProVal1yLeuG1yG1yG1uA1a 2100
Db      6241 CACCTGAGGGGAGAGCTGCGGCCCAAGAGCCAGCCCGCTGGAAGCTTGGGGGAGGCC 6300
Qy      2101 A1aH1aLeuProH1aLeuAAspLyAG1uProLyG1uSerG1uProSerSerSerProLeu 2120
Db      6301 GCCCAGCTCCACACCTGCGGCCCTGCTGAGAGCAGCCCTGCTGACAGCCCTGCTC 6360
Qy      2121 G1uThrA1aProG1yVal1yLyAG1yH1aG1uAAspLyA1aThrLeuA1aG1uH1aSer 2140
Db      6361 CAGACCCGCCCGAGGGGTCAAGAGTCAACGCGGGTGTCACTCGGCCACACATCAGT 6420
Qy      2141 G1uVal1LeuThrG1uAAspLyTThrArgH1aH1aProG1uG1uLeuSerA1aProLeuPro 2160
Db      6421 GAGGTCA1aCAGACGACTACACCCGGCACCACCAAGAGCTCAAGCGACCCCTGCCC 6480
Qy      2161 A1aProLeuTySerPheProG1yA1aSerCyBProVal1aLeuAAspLyAAspLyProPro 2180
Db      6481 GCCCCTCTACTCTCTCCCTGGGGCAGCTGCCGCTGAGACTCCGCGCCGACACC 6540
Qy      2181 SerAAspLeuTyLeuProProProProAAspH1aG1yA1aProA1aArgG1ySerProH1aSer 2200
Db      6541 AGTACCTTACCTCCGCGCCCGGACCATGTGCCCCGCGCTGCTCCCGCCACAGC 6600
Qy      2201 G1uG1yG1yLyAAspSerProG1uProAAspLyH1aSerVal1aLeuG1yG1yG1yG1yAAsp 2220
Db      6601 GAAGGGGGCAAGAGTCTCCAGAGCCAAACAGACGCTGCTTGGGTGGTGGTGGAGAC 6660
Qy      2221 G1yT1aG1uProVal1aSerProProG1uG1yMetThrG1uProG1yH1aSerArgSerA1a 2240
Db      6661 GGTATTGAACCTGCTGCCACCGGAGGATGACGAGCAGGAGGAGGAGCTCCGGAGTGCT 6720
Qy      2241 ValTyProLeuLeuTyArgAAspLyG1uG1uThrG1uProSerAAspMetG1ySerLyS 2260
Db      6721 GTGTACCCGCTGCTGACCGGAGTGGGAAACAGCGAGGCCACAGAGATGGCTCCAG 6780
Qy      2261 SerProG1yAAspThSerG1uProProA1aPhePheSerLyLeuThrG1uSerAAspSer 2280
Db      6781 TCTTCAGGCAACACAGCCAGCCGCTCTTTCAGCAAGCTGACGAGAGCACTCC 6840
Qy      2281 A1aMetVal1ySerLyS1yG1uG1uAAspLyA1aSerLyLeuAAspThH1aAAspArgAsn 2300
Db      6841 GCCATGGTCAAGTCAAGAGCAAGAGATCAACAAAGAGCTGAACACCAACCGGAAAT 6900
Qy      2301 G1uProG1uTyArgH1aSerG1uProG1yThrG1uLeuPheAAspMetProA1a1LeuThr 2320
Db      6901 GAGCTGGAATATACATATACAGCCCGGAGCGAGTCAATGCTTCAATAGCCCGCCATCAC 6960
Qy      2321 G1yThG1yLeuMetThTyArgSerG1uA1aValG1uG1uH1a1aSerThH1aAspMet 2340
Db      6961 GGAACAGGCTTATGACTTATAGAGCCAGCGGTGAGAACTGCGACAGCAACACATG 7020
Qy      2341 G1yLeuG1uA1a1e1aArgLyA1aLeuMetG1yLyTyArgG1uProG1uG1uSer 2360
Db      7021 GGCTGAGAGGCAATATTAGAAAGGCACTCATGGTAAATATAGCAGTGGAAAGAGTCC 7080
Qy      2361 ProProLeuSerA1aAAspA1aPheAAspProLeuAAspA1aSerA1aSerLeuProA1aA1a 2380
Db      7081 CCGCCGCTGAGCGCAATGCTTTAACTCTGAAAGCCAGTGCACACCTGCGCGCTGCT 7140
Qy      2381 MetPro1LeuThH1aA1aAAspG1yArgSerAAspH1aThLeuThH1aSerProG1yG1y 2400
Db      7141 ATGCCCATTAACCGCTGCTGACGAGAGTGAACACACATCACTGCGCAGGTGGGGC 7200
Qy      2401 G1yLyA1aA1yValSerG1yArgProSerSerArgLyA1aA1ySerProA1aProG1y 2420

```

```

Db      7201 GGGAGGCGCAAGGCTCTGACAGACCCAGACCGGAAAGCAAGTCCCGGCCCGGCG 7260
Qy      2421 LeuA1aSerG1yAAspAAspProProSerVal1aSerSerVal1aSerG1uG1yAAspCyAAsn 2440
Db      7261 CTGGCATCTGAGGACCGGCCACCTCTGTCTCTCAAGTGAACCTGAGAGGAGACTGCAAC 7320
Qy      2441 ArgH1aThProLeuThH1aAAspVal1yProG1uAAspAAspProSerSerA1aG1ySerThr 2460
Db      7321 CGCGGAGCGCGCTCAACCAACCGGTGTGGAGAGACAGGCGCTGCTCGGAGGTTCCACG 7380
Qy      2461 ProPheProTyArgProLeuLeu1eMetArgLeuG1uA1aG1yVal1aMetA1aSerProPro 2480
Db      7381 CCAATCCCTTACCAACCCCTGATCATGCGCTGACAGCGGGGTGTCATGCTTCCCAACC 7440
Qy      2481 ProProG1yLeuProA1aG1ySerG1yProLeuA1aG1yProH1aH1aA1aTPAAspG1u 2500
Db      7441 CACCGGGCTCCCGGGGAGGGGCGCCCTGCTGAGCCCAACAGCCTGGAGACAG 7500
Qy      2501 G1uProLyAAspProLeuLeuCySerG1uTyArg1uThH1aSerAAspSerG1u 2517
Db      7501 GAGCCAGGCACTGCTGCTGCTGCGAGTACGAGACACTTCCAGACGAG 7551

RESULT 2
ADG86298 standard; cDNA; 8561 BP.
ID ADG86298
XX AC
XX ADG86298;
XX AC
XX 11-MAR-2004 (first entry)
XX DE
XX Human SMRT encoding cDNA SEQ ID NO:12.
XX SMRT: silencing mediator for retinoid and thyroid hormone action;
XX SMRT inhibitor; cytosolic; antiinflammatory; antiarthritic;
XX antirheumatic; antisense therapy; inflammatory disorder;
XX rheumatoid arthritis; hyperproliferative disorder; cancer; leukaemia;
XX breast cancer; human; gene; ss.
XX OS
XX Homo sapiens.
XX FH
XX Key 2. 7555
XX CDS /*tag= a
XX FT /product= "SMRT"
XX PN
XX NO2003106645-A2.
XX PD
XX 24-DEC-2003.
XX PF
XX 17-JUN-2003; 2003MO-US018923.
XX PR
XX 17-JUN-2002; 2002US-00174014.
XX PA
XX (ISIS-) ISIS PHARM INC.
XX PI
XX Bennett CF, Freier SM, Doble KM;
XX WPI; 2004-082184/08.
XX DR
XX P-PDB; ADG86298.
XX DR
XX GENBANK; NM_006312.
XX PT
XX Novel antisense compound targeted to nucleic acid encoding SMRT
XX (silencing mediator for retinoid and thyroid hormone action), useful for
XX treating animal having disease associated with SMRT such as cancer,
XX rheumatoid arthritis.
XX PT
XX Example 15; SEQ ID NO 12; 260bp; English.
XX PS
XX The present invention describes a compound (I) 8-50 nucleobases in length
XX targeted to a nucleic acid molecule encoding SMRT (silencing mediator for
XX retinoid and thyroid hormone action), where (I) specifically hybridises
XX with the nucleic acid molecule encoding SMRT and inhibits expression of a
XX SMRT. (I) specifically hybridises with at least 8-nucleobase portion of a

```

CC preferred target region on nucleic acid molecule encoding SMRT. Also described is a composition (II) comprising (I) and a carrier or diluent. (I) and (II) have cytostatic, antiinflammatory, antiarthritic and CC antirheumatic activities, and can be used in antisense therapy, and as SMRT expression inhibitors. (I) is useful for inhibiting the expression of SMRT in cells or tissues. (I) is also useful for treating an animal having a disease or condition associated with SMRT, e.g., inflammatory disorder such as rheumatoid arthritis, or a hyperproliferative disorder such as cancer chosen from leukaemia and breast cancer, by inhibiting the expression of SMRT. (I) is useful for diagnostics, therapeutics, prophylaxis and as research reagents and kits. The present sequence CC encodes human SMRT, which is used in an example from the present CC invention. N.B. The present sequence is designated as SEQ ID NO:11 in CC example 15 but corresponds to SEQ ID NO:12 in the Sequence Listing.

XX Sequence 8561 BP; 1862 A; 3033 C; 2525 G; 1141 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	8561
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-09-522-753-5 (1-2517) x ADG86298 (1-8561)

```

QY 1 MetSerG1SerThG1nLeuVal1aG1nThTPArG1aThnG1uProArG1YrPro 20
Db 2 AAGTCGAGGCTCCACACAGCTTGtGSCAGACGTGAGGAGCCACtAGCCCGCTACCG 61
QY 21 ProHisSerLeuSerTyRProValG1n1ealAarG1ThiStrApyA1G1yLeuLeu 40
Db 62 CCCACAGGCTTTCtACCAAGTCAGATCGCCGAGCCAGCAACGAGCGTGGGCTCTCG 121
QY 41 GluTYrG1nHisSerAArgApyTYrAlaSerHisLeuSerProG1ySer1le1egin 60
Db 122 GAGTACACACACACACCTCCGCGACTAGCTGCTCCACCTGTCCGCGGCTCCATCACCAG 181
QY 61 ProG1nArG1ArG1rProSerLeuSerG1uPhG1nProG1YanG1uArG1SerG1n 80
Db 182 CCCACGCGGAGAGGCTCTCTCTGCTGAGTTCAGACCCCGGAAATAGACGATCCAG 241
QY 81 GluLeuHisLeuArG1ProG1uSerHisSerTYrLeuProG1uLeuG1YySerG1uMet 100
Db 242 GAGCTCACTCGGCGGCGAGTCCCACTACCTGCCGAGCTGGGAGTCAAGTCAAGATG 301
QY 101 GluPhe11eG1uSerTySArG1rProArG1uLeuG1uLeuProArG1ProleuLeuArG1Pro 120
Db 302 GAGTTCATGAAAGCAAGGCGCTCCGCTAGAGCTCTGCTGACCCCTGCTGCACCG 361
QY 121 SerProLeuLeuAlaThrG1G1nProAlaG1ySerG1uAsPLeuThr1ySArG1rSer 140
Db 362 TCACTCTCTGCGCGGAGGAGCTCGGAGATCTGAAGACTTCAACAGAGCCATTAG 421
QY 141 LeuThrG1YyLeuG1uProValSerProProSerProProHisThrAAsProG1uLeu 160
Db 422 CTGACGGGAGAGTGGAAACGGTGTCTCCCGGAGCCCGGCACTGACCTGAGCTG 481
QY 161 GluLeuVal1rProProArG1uSerTyS1yG1uG1uLeu11eG1nAsMetAsPArG1V1aAsP 180
Db 482 GAGCTGAGTCCGCGACGCGCTGTCCAAGGAGAGCTATCCAGAACTAGGACCCCGTGGAC 541
QY 181 ArG1G1u11eThrMetVal1G1uG1nG1n1eSerTyS1yLeu1yS1yG1nG1nG1n1e 200
Db 542 CGAGAAATCAACATGGTGAAGCAGAGATCTTAAGCTGAAGAAAGAGCAGCAACGCTG 601
QY 201 GluG1uG1uAla1a1a1yS1rProG1uProG1u1yS1rProValSerProPro11eG1u 220
Db 602 GAGGAGAGAGGCTGCCAAGCGCGCCGAGCTGAGAGACCGGCTGCACCGCGCCATCGAG 661
QY 221 SerTyS1yHisArG1SerLeuValG1n11e11eTYrAsP1uAsnArG1YyS1yS1a1aG1u1a 240

```

```

Db 662 TCGAAGACCCGAGCTGGTGCAGATCATTCAGACGAGAACCCGAGAAAGCTGAAGCT 721
QY 241 AlaHisArG11eLeuG1uG1yLeuG1YrProG1nVal1G1uLeuProLeuTyRanG1nPro 260
Db 722 GCACATCCGATTCTGGAAGGCTGGGGGCCAGG1yGAGCTCCGCTGTACAAACAGCC 781
QY 261 SerAsPThrArG1nTYrHisG1uAsn11e1yS11eAnG1n1a1MeArG1YyS1yS1e 280
Db 782 TCCGACACCCCGCACTATCATAGAAACATCAAAATMAACAGAGCATGGGAGAAAGCTA 841
QY 281 11eLeuTYrPhLeuArG1rAAsnHis1a1aArG1yS1nTPrTyS1yS1yPhCyS1n 300
Db 842 ATCTTG1ACTTAAGAGAGAAATACGCTCCGAAACATAGAGAGAGATTTCTCCAG 901
QY 301 ArG1TYrAsP1uLeuMetG1uAlaLeuG1u1yS1yS1a1G1uArG11eG1uAsnAsnPro 320
Db 902 CGCTATGACCACTCATGAGAGCTTGGAAAAGGTGAGCGCATCGAAAACACCCG 961
QY 321 ArG1ArG1rG1a1a1yG1uSerTyS1yValArG1uTYrTYrG1u1yG1nPhrProG1u1e 340
Db 962 CGCCGCGCGCCAGAGAGAGAGTGGCGGAGTACTACGAAAGCATGTTCCCTGAGATC 1021
QY 341 ArG1yS1G1nArG1uLeuG1nG1uArG1MetG1nSerArG1ValG1YnArG1ySerG1y 360
Db 1022 CGCAAGCAGCGGAGCTGCAGAGCGCATGCAAGCGAGGTGGCCAGCCGCGGAGTGG 1081
QY 361 LeuSerMetSerAla1a1aArG1SerG1uHisG1uVal1SerG1u11eAsP1yLeuSer 380
Db 1082 CTGTCTCATGTGGCGCGCGCGAGCGAGCAAGG1yTGCAGATCATGATGAGCTCTCA 1141
QY 381 GluG1nG1uAsnLeuG1u1yS1e1MeArG1nLeu1a1a11eProProMetLeuTYr 400
Db 1142 GAGCAGAGAGAACTGAGAGAGAGATGCGCCAGCTGGCGGTATCCCGCCATGCTGATC 1201
QY 401 AsP1aAsP1nG1nArG11e1yS1rPhe11eAsnMetAsnG1YyLeuMet1aAsP1rMet 420
Db 1202 GAGGCTGACACAGCGCATCAAGTTCATCAATGAAAGGCTTATGGCCACCCAG 1261
QY 421 TyS1ValTYrTyS1AsPArG1nVal1MeArMetTrP1SerG1uG1nG1u1yS1G1uThrPhe 440
Db 1262 AAGGTGTAAAGACCGGACAGGTCAAGAACTTGTGAGTATGATGATCTCTGAGAG 1321
QY 441 ArG1u1yS1rPheMetG1nHisProTyS1AsnPhrG1YyLeu11eAlaSerPheLeuG1uArG 460
Db 1322 CGGAGAACTTCATGACAGATCCCAAGAACTTGTGAGTATGATGATCTCTGAGAG 1381
QY 461 TyS1ThrVal1aG1uCyS1eVal1e1uTYrTYrTYrLeuThrTyS1yS1eAsnG1uAsnTYrTyS 480
Db 1382 AAGACAGTGGCTGAGTGCCTCTTAATTACTACCTGACTTAAGAAATAGAACTTAAG 1441
QY 481 SerLeuValArG1rArG1SerTYrArG1rArG1YyS1SerG1nG1nG1nG1nG1n 500
Db 1442 AGCTGTGTAGAGGAGGCTATGCGCGCGCGCAAGAGCAGACAGCAACAGCAGCAG 1501
QY 501 G1nG1nG1nG1nG1nG1nG1nG1nG1nProMetProArG1SerG1nG1uG1u1yS 520
Db 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCCATGCCCCGAGCAGCAGCAGCAGGAGAA 1561
QY 521 AsP1u1yS1yG1u1yS1yG1u1aG1u1aG1u1yS1yG1u1yS1rProG1uVal1G1uAsn 540
Db 1562 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
QY 541 AsP1yS1G1uAsP1uLeu1yS1u1yS1rThrAsPArThrSerG1Yy1uAsPAsnAsP1u 560
Db 1622 GACAGAGAAAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 TyS1G1u1a1aVal1aSerTyS1yArG1yS1rThr1aAsnSerG1nG1YArG1rG1yS1y 580
Db 1682 AAGGAGGCTGTGCTCCAAAGCGCCGCAAAACTGCCAAACAGCAGGAGAGAGAGAGAG 1741
QY 581 ArG11eThrArG1SerMet1a1aAsnG1u1aAsnSerG1uG1uAla11eThrProG1nG1n 600
Db 1742 CGCATCAACCGCTCAATGAGCTAATGAGCCCAACAGAGAGAGGCTATCAACCCCGAGCAG 1801

```

Qy	601	SeraIaGlueuaIaSerMetGlueuaangIuSerSerAgyTrpThGlueIuGlueMet	620
Db	1802	AGCGCCGAGCTGGCTCATGAGAGCTGAATGAGATTCTCCCTGGACAGAAAGAAATG	1861
Qy	621	GIuThraIaIyIySgIyLeuLeuGIuHiaGIyArgAntIpsSerAIIaIaIArgMet	640
Db	1862	GAACACGCAAGAAAGAGTCTCTCGAACAAGCCGCAACTGATGCGCATGCCCGGATG	1921
Qy	641	ValGIySerIySthVaISerGIuCyIaIySaAnPheTyrPheAsnTyCIyIySaArgGIu	660
Db	1922	GTGGGCTCCAAAGACTGTCTCCCAAGTAAAGAACTTCTACTTCAACTAACAAGAAGCGAG	1981
Qy	661	AsnLeuAaBGIuIIeLeuGIuHiaIySleuIyMetGIuIySgIuAaSaIaArg	680
Db	1982	AACTCGATGAGATCTTGCACACAGCAACAAGTGAAGAGAAAGAGAAAGACGCCGG	2041
Qy	681	ArgIySlySlySaIaProIaIaIaISerGIuIaIaIaPheProProValIaI	700
Db	2042	AGAAAGAAAGAAAGCCGCGCGCGCGCACGAGGAGGCGTGCATTCGCCCGGTGTG	2101
Qy	701	GIuAaBGIuIuMetGIuIaISerGIyValISerGIyaangIuGIuIuMetValGIuGIu	720
Db	2102	GAGATGAGAGATGAGAGCGCTCGGCGCTGACCGGAATGACAGAGATGCTGAAGAG	2161
Qy	721	AlaGIuIaIeUHiIaIaSerGIyAaSGIuValProArgIyGIuCySerGIyProIa	740
Db	2162	GCTAAGCCTTAACATGCTCTGTGGAAATAGGTGCCCAAGAGGGAATGACATGGCCAGCC	2221
Qy	741	ThrValAaSaSaSerSerAaPThrGIuSerIleProSerProHiaSthGIuIaIaIyS	760
Db	2222	ACTGTCAACAACAGTCTACAGACACGAGAGCATCCCTCTCTCACTCAAGAGCGCAAG	2281
Qy	761	AaPThrGIyIaAaSGIyProIySaProProIaThrIeUgiYAlaAaBGIyProProPro	780
Db	2282	GACACAGGAGCAAGATGGGCCCAAGGCCCAACCAACCTGGGGCGCGAAGGSCACCCCA	2341
Qy	781	GIyProProThrProProIaArgIyThSerArgAlaPoiIeGIuProThrProIaSer	800
Db	2342	GGCCACCCCAACCCACCAACGAGAGACATCCGGGCCCCCATTTGAGCCACCCCGCTCT	2401
Qy	801	GIuIaIaThrGIyAlaProThrProProProIaIaProProSerProSerAlaProProPro	820
Db	2402	GAAGCCACCGAGCCCTTACGCCCCCCACACACACCCCATGCCCCCTTGACCTCTCTCT	2461
Qy	821	ValValProIySgIuIySgIuGIuIuThraIaIaIaIaProProValGIuGIuGIy	840
Db	2462	GTGGTCCCCAAGAGAGAGAGAGAGAGACCGCAGCAGCGCCCCCACTGGAAGAGGGG	2521
Qy	841	GIuGIuGIuIyProProIaIaIaGIuIuIeUaIaValAaPThrGIyIySaIeGIuGIu	860
Db	2522	GAGAGAGCAAGACCCCCCGCGCTGAGAGAGCTGCAGTGCACACAGGGAAGGCGCAAGAG	2581
Qy	861	ProValIySsSerGIuCyThSthGIuGIuIaIaGIuGIuGIyProIaIySgIyIySaAaIa	880
Db	2582	CCCGTCAAGACCGAGTGCACCGAAGAAACCGAAGAGGGGCCGCAAGGCGCAAGACGCG	2641
Qy	881	GIuIaIaIaGIuIaIaThraIaIaGIuGIyAlaIeUySaIaGIuIySgIuIyGIySer	900
Db	2642	GAGGCCGCTGAGGCCACGCGCGAGGGGGCGCTCAAGAGCAGAGAGAGAGAGAGGGGAGC	2701
Qy	901	GIyIyArgIaIaThrThraIaIySsSerGIyIaIaProGIuIaAaPSerAaPSerSerAlaThr	920
Db	2702	GGCAGGGCGCACACTGCTCCMAAGTCTGGGCGCCCCCAAGACACAGCACTCCAGTGCATCC	2761
Qy	921	CySerSerIaAaBGIuValaAaBGIuAlaGIuGIyGIyAaPlySaAaIyGIuIeUySerPro	940
Db	2762	TGCAGTGCAGACGAGTGCATGAAGGCCAGAGGCCGCGCAACAAGACCGGCTCTCTCCCA	2821
Qy	941	ArgProSerIeUyThrProThrGIyAaPProIaIaAaIaSerProGIuIyPro	960
Db	2822	AGGCCACGACCTCTCAACCCCGCATGCGACCCCGGGCGCAATGCTCAACCCCAAGAACCA	2881

QY	961	LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLys	980
Db	2882	CTGAACCTGAAAGCAGCTGAAAGACAGCGGCTGCATCCCCCATTCACAGTCAACAA	2941
QY	981	ValHisGluProProAspArgGluAspAlaAlaProThrLysProAlaProProAla	1000
Db	2942	GTCCATAGACCCCCCGGAGAGAGACACTCCACCAAGCCAGTCCCCCGAGCCACCG	3001
QY	1001	ProProGlnAenLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg	1020
Db	3002	CCACCGCAAAACCTGCAGCGCGGAGAGGAGCCCTCAGCAGCCTGGGACGACCCCGG	3061
QY	1021	GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln	1040
Db	3062	GGCAGAGACAGAGACCCCGGCACCCCGCCGACAAAGAGGCTTCGACGCGGAGGCCAG	3121
QY	1041	LysLeuProGlyAspProProProCysThrThrSerGlyLeuProPheProValProProArg	1066
Db	3122	AAGCTGCTGGGGACCCCCCTTGCTGGACTTCGGGCTGGCCCTTCGCCGATCCCCCGT	3181
QY	1061	GluValAlaIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro	1080
Db	3182	GAGGTGATCAAGGCTCCCGGCAGTCCCGGACCCCTCAGCTTCTCTAGCTCCACCT	3241
QY	1081	GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro	1100
Db	3242	GGTCAACCACTGCCCCCTGGGCTCCTAAGACACTCCGGCCCGTCTGGCCGCGCCACCC	3301
QY	1101	ThrIleSerAenProProProProIleIleSerSerAlaLysHisProSerValLeuGluArg	1120
Db	3302	ACCATCTCAACCCGCGCTCCCTCATCTCTCTGCCAAAGCACCCACAGCTCTGAAGAG	3361
QY	1121	GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlnHis	1140
Db	3362	CAATTAAGTGGCATCTCCCAAGAAATGTCGTCACGCTCAACGTCCTGACTACAGCAT	3421
QY	1141	AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	3422	GCCAAAGCCCCGGGGGCGCTGTGCACATGGGGCTGCCCTGACCATAGACCCCAAAAG	3481
QY	1161	LeuAlaProPheSerArgLysValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro	1180
Db	3482	CTGGCACCTTCAGCGGAGTGAAACAGAGACAGCTGTCCCAACGGGGCCAGGCTGGCCCA	3541
QY	1181	ProGluSerLeuGlyValProThrAlaGlnGluLysSerValLeuAspGlyThrAlaLeu	1200
Db	3542	CCGAGAGAGCTGGGGGGTGGCCACAGCCACAGAGGCGTCTGCTGAGAGGGACAGCTCTG	3601
QY	1201	GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp	1220
Db	3602	GGCTCAGTCCGGGCGGAGAGATCATCCAAAGGATTCACAGCACAGGGTCCCTCGGAC	3661
QY	1221	SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLys	1240
Db	3662	AGCCCATCATCATCCCGGCTCATCACCCACGCGCAGCGCATGACGTCTGTGTAAAG	3721
QY	1241	GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu	1260
Db	3722	GGCACCATACACAGATCATGGGAGAGACGCCCGAGTGGCTTGGACCGCGCGGGAG	3781
QY	1261	AspSerLeuProLysGlyHisValIleTyrGlnGlyLysLysGlyHisValLeuSerTyr	1280
Db	3782	GACAGCTGGCCMAAGGCGCAAGTATCTTAGAAGGCAAGAGGCGCAAGTCTTGTCTAT	3841
QY	1281	GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro	1300
Db	3842	GAGGTGTGCATGTCTTGACCCAGTGTCTCCAAAGAGACGCGCAAGACACTCAGGACCC	3901
QY	1301	ProHisGlnThrAlaAlaProLysArgTyrThrAspMetMetGluArgValAlaGlyArg	1320
Db	3902	CCCATATGACGGCGCGCCCAAGCGCATTAATGACATGATGGAGGGCCCGCTGGGACGA	3961
QY	1321	AlaIleSerSerAlaSerIleGlnGlyLeuMetGlyArgAlaIleProProGluArgHis	1340

|||||
Db 3962 GCCATCTCTCAAGCAGCATCAAGGTCTCATGGGCGTGCATCCCGCGGAGCGACAC 4021
QY 1341 SerProHisHisLeuLeuysGluGlnHisHisGlySerIleThrGlnGlyIlePro 1360
Db 4022 ACCCCCCACACCTCTCAAGAGCAGCACACATCCGGGGTCCATCCACACAGGGAATCCCT 4081
QY 1361 AGSerTYrValGluValGlnGluAspTYrLeuArgArgGluValAlaLeuLeuLysArg 1380
Db 4082 CGGTCTTACGTGAGGAGCAGAGAGACTACCTGCGTGGGAGGCCAAGCTCTCAAGCGG 4141
QY 1381 GluGlyTYrProProProProProSerArgAspLeuThrGluAlaTYrIleThrGln 1400
Db 4142 GAGGGGACCGCTCCGCCCCACCGCCTCAAGGAACTGACCGAGGCTCAAGAGCGCAG 4201
QY 1401 AlaLeuGlyProLeuLeuLeuLysProAlaHisGlnGlyLeuValAlaThrValLysGlu 1420
Db 4202 GCCCTGGGGCCCCCTGAAGCTGAAGCGGGCCCATGAAGGCGCTGGTGGCCACGGTGAAGAG 4261
QY 1421 AlaGlyArgSerIleHisGlnIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db 4262 GGGGGCGGCTCATCATAGATCCCGCGAGAGACTGGGAGCACGCCCGAGCTGCC 4321
QY 1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyTYrProLeuLysTYrAsp 1460
Db 4322 CTGGCCCCCGGCGCTCAAGGAGGGGCTCCATCAAGGAGCGACCCCGCTCAAGTACGAC 4381
QY 1461 ThrGlyAlaSerThrThrGlySerIleLysHisAspValArgSerLeuIleGlySerPro 1480
Db 4382 ACCGGGCGCTCCACCTGCTCCAAAAGACAGAGAGCTGCTCCCTCATCGGACACCCC 4441
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db 4442 GGGCGGACGTTCCACCGCTGCACCCGCTGGATGTATGGCCGACCCCGGGAGCTGGAA 4501
QY 1501 ArgAlaCysTYrGluGlnSerLeuLysSerArgProGlyTYrAlaSerSerSerGlyGly 1520
Db 4502 CGTGGCTGTCTCAAGAGAGAGCTGAAGAGCGGCGCAGGAGCGCAGCAGCTGGGGGGC 4561
QY 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro 1540
Db 4562 TCCATTGCGCGGGCGGCCCGCGCTGATGTGCTGACCTGGTAAAGCCGGCGAGACCCC 4621
QY 1541 LeuThrTYrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db 4622 CTGACCTATGAGGAGCACCGGGGACCCCTTTGCCGGCACCTCCACGAGGTTGGCCGCG 4681
QY 1561 ThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerIleAspSer 1580
Db 4682 ACCATGCGGAGGCCACGCGCGGCTGCAGAGGAGGAGCTTTGCTCCAGCAGGCATCC 4741
QY 1581 GlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrVal 1600
Db 4742 CAGGACCGAAAGCTGACGTGAGCTCTGAGATGCGCAAGTCCCGCGCACACACCGTG 4801
QY 1601 ProGluHisHisProHisProIleSerProTYrGluHisLeuLeuArgGlyValSerGly 1620
Db 4802 CCGGAGACACACACACACCCCATCTGCGCTTATGAGCAGCTGCTTGGGGGCGTGAATGGC 4861
QY 1621 ValAspLeuTYrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 4862 GTGGACCTGTATGAGCAGCACATCCCGCTGGCTTGACCCGACCTCCATACCCCGCGG 4921
QY 1641 IleProLeuAspAlaAlaAlaAlaTYrTYrLeuProArgHisLeuAlaProAspProThr 1660
Db 4922 ATCCCTCTGGACCGACCGCTGCTTACTACTGCCCCGACACCTGGCCCCCAACCCACCC 4981
QY 1661 TYrProHisLeuTYrProProTYrLeuIleArgGlyTYrProAspThrAlaAlaLeuGlu 1680
Db 4982 TACCCGACCTGTACCAACCTTACTCATCGGGCTACCCGACGAGGGGGCGCTGGAG 5041
QY 1681 AsnArgGlnThrIleIleLeuAspTYrIleThrSerGlnGlnMetHisHisAsnThrAla 1700
|||||

Db 5042 AACGGCAGACCATCATCATAGTACTACCTCGCAGACAGATGCACCAACAGGGCC 5101
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerLeu 1720
Db 5102 ACCGCGATGGCCAGAGGACTATATGCTGAGGGGCTCTGCCCGCGAGTCTTGCTG 5161
QY 1721 AlaLeuAsnTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
Db 5162 GCATCAACTAGCTGGCGGGTCCCGAGGACATCATGACTGTCTCCMACTGCACACCTG 5221
QY 1741 ProValLeuValProProThrProGlyTYrProAlaThrAlaMetAspArgLeuAlaTYr 1760
Db 5222 CTGTGCTGTGCCCCCGACACCAAGCACCACCGCACCGCATGGACCGCTTGCTG 5281
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db 5282 CTCCCAACGGGCCCCAGCGCTTCAGAGACGGCCACAGAGCTCCCACTTCCACGAG 5341
QY 1781 GlyProThrHisLeuThrTYrAspProThrThrThrSerSerSerGluArgGluArgAspArg 1800
Db 5342 GGTCCAAACACCTTGCAAAACCAACACACAGTCTCTCGAGGGGAGGAGACCGG 5401
QY 1801 AspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThrThr 1820
Db 5402 GATCGAGAGCGGAGCCGGGATCGAGCCGGAAGAGTCCATCTTCACTGCACACGAGAG 5461
QY 1821 ValGluHisAlaProIleTYrArgProGlyTYrThrGlnGlnSerSerGlySerSerGlySer 1840
Db 5462 GTGGAGACAGCACCATCTCGAGACTGGTACAGAGACAGCAGCGGAGCGGACG 5521
QY 1841 SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5522 ACCGGGGGGGGGGGGGAGCAGACCGCCCGGCTCCCATCTCCATCCACCCACGAC 5581
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsn 1880
Db 5582 TGCCCATCTCCCTCGAGCCAGAGATGCTTCCAGACAGACCCAGATGCTTCAAC 5641
QY 1881 ThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSer 1900
Db 5642 ACGGATAGAGGATATCATACCGCTGTGAGGCCAGAGCCAGACCGCATCGTGAAGTCC 5701
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
Db 5702 ACTCCACTCTCTCAACCGCTTGCCAGCTGCACATTCCTCCACTGCGACCACTGCCA 5761
QY 1921 LeuGlyGlyTYrHisLeuAspGlyValTYrProThrLeuMetGluProValLeuLeuProLys 1940
Db 5762 CTGGGGGAGACCTCGATGGGGGTACCTTACCTCATGAGCCGCTTGCTGCCCCAAG 5821
QY 1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
Db 5822 GAGGCCCCCGGGTGGCCCGGGCCAGAGCGGCCCGAGCAGACACCGGCATGCTTCTC 5881
QY 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db 5882 GCCAAGCCCCCAGCCCGCTCGGGCTGAGACCGCTCTCCCGCCGACAAAGGCTCTGGAG 5941
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys 2000
Db 5942 CCCCCGCTGTAGTGTCTCTGTGTGGCACGGCCACATGCGCCGACCCCTGGGAAG 6001
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
Db 6002 AACCTCGACCTTCACACCGCCAGCCGAGCCGCGCGGACCTGCTCGGCTGGAG 6061
QY 2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSer 2040
Db 6062 CGGACCGGGAAGAAAGCTAAAGTAAACCTTTTCATGCAAGAACTGGAACCTCGTTCT 6121
QY 2041 LeuGlyTYrHisGlySerSerTYrSerProGluGlyValGluProValSerProValSer 2060
Db 6122 CTGGGTTACAGGGCAGCACTAACGCCGCCGAAGGGGTGAAGCCGTGAGCCCTGTGAGC 6181

QY 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer 2080
 DB 6182 TCACCCAGTCTGACCCAGCAAGAGGGGTCCCAAGACCTCGAAGAGCTGACAGAGGC 6241
 QY 2081 HisLeuGluGluGlyLeuLeuArgProLysGluInProGlyProValLysLeuGluGlyGluA 2100
 DB 6242 CACCTGAGGGGGAGCTGCGGGCCAGACAGCCCGGCTTGGAGCTTGGGGGGAGGGCC 6301
 QY 2101 AHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
 DB 6302 GCCCACCCTCCACACCTGCGGGCCCTGCTGAGAGCCAGCCCTGCTCAGCCCTGCTGTC 6361
 QY 2121 GlnThrLysProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisLysSer 2140
 DB 6362 CAGACCGCCCGCAGGGGCTCAAGGTCACAGCGGGTGTGACCTGGCCAGCAGCATCAGT 6421
 QY 2141 GluValLysThrGlnAspLysThrArgHisHisProGlnGlnLeuSerAlaProLeuPro 2160
 DB 6422 GAGGTCAATCACACGAGCTACACCGGACACCCAGCAGCAGCTCAGCGCAGCCCTGCC 6481
 QY 2161 AlaProLeuLysSerPheProGlyAlaSerCysProValLeuAspLysArgProPro 2180
 DB 6482 GCCCCTCTTACTCTCTCTGCGGGCCAGCTGCCCCCTGCTGAGCTCCGCGCCACACC 6541
 QY 2181 SerAspLeuLysLeuProProProProAspHisGlyValAlaProAlaArgGlySerProHisSer 2200
 DB 6542 AGTGACCTTACCTCCGCGCCCGGAGCAGTGTGTCGCCGCGCTGCTCCCGCCACAGC 6601
 QY 2201 GlnGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGluGlyGlyGluAsp 2220
 DB 6602 GAAGGGGGCAAGGCTCTCCAGAGCCAAACAGACGTCGGTCTTGGTGGTGGAGAGAC 6661
 QY 2221 GlyLysGluProAlaSerProProGluGluGlyMetThrLysProGlyHisSerArgSerAla 2240
 DB 6662 GGATATTGAACCTGTGTCCTCCAGAGGAGGATGACGAGGACGAGGCTCCCGAGTGTCT 6721
 QY 2241 ValLysProLeuLeuLysArgAspGlyGluGlnThrGluProSerArgMetGlySerLys 2260
 DB 6722 GTGTACCCGCTGCTGTCGCGGATGGGAAACAGCGAGCCAGCAGAGATGGCTCCAG 6781
 QY 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
 DB 6782 TCTCCAGGCAACACGAGCCGCGCAGCTTCTTCAAGAGCTGACCGAGAGCAATCC 6841
 QY 2281 AlaMetValLysSerLysLysGlnGluLeuAsnLysLysLeuMetThrHisAsnArgAsn 2300
 DB 6842 GCCATGCTCAAGTCCAGAGGAGCAAGATCAACAGAGAGCTGAACACCCACACCGGAAT 6901
 QY 2301 GluProGluLysAsnLysSerGlnProGlyThrGluLeuPheAsnMetProAlaLysThr 2320
 DB 6902 GACCTGGAATACATATACAGCCAGCTGGAGCGGAGTCTTCAATATGCCCCCATACCC 6961
 QY 2321 GlyThrGlyLeuMetThrLysArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet 2340
 DB 6962 GGAACAGGCTTATGACTATAGAACGACGCGGTGACGAGAACTGCACACCAACATG 7021
 QY 2341 GlyLeuGlnAlaLysLeuArgLysAlaLeuMetGlyLysLysArgGlnProGluGluSer 2360
 DB 7022 GGGCTGAGGCTCAATATTAGAAAGGACCTCATGTGTAATATGACCGTGGAGAGATCC 7081
 QY 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
 DB 7082 CCGCCCTCAGCCGCCAATGCTTTTAACTCTGAATCCAGTCCAGCTGCGCGCTGCT 7141
 QY 2381 MetProLysThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
 DB 7142 ATCCCATATACCCCTGCTGACGAGCGAGTGACACACACTCTGCGCAGGTGGCGGC 7201
 QY 2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysValLysSerProAlaProGly 2420
 DB 7202 GGGAGGCGCAAGGCTCTGGAGAGCCAGCAGCGGAAAGCAAGTCCCGCGCCCGGGC 7261

QY 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsn 2440
 DB 7262 CTGGCATCTGGGAGCCGCGCACCTTGTCTCTCAATGACCTCGAGGGAGACTGCCAAC 7321
 QY 2441 ArgArgThrProLeuThrAsnArgValTTPGluAspArgProSerSerAlaGlySerThr 2460
 DB 7322 CGCGGAGCGGCTCAGCAACCCGCTGTGGAGGACAGGCTGCTGCGCAGGTTCCAGC 7381
 QY 2461 ProPheProLysProLeuLysMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
 DB 7382 CCAATCCCTTACAAACCCCTGATCATCGGCTGCAAGCGGGTGTATGCTTCCCAACC 7441
 QY 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaLysProGlu 2500
 DB 7442 CCAACCGGCGCTCCCGCGGAGGAGGCGCCCTGCTGCGCCCGCCACAGCCTGGGAGCAG 7501
 QY 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
 DB 7502 GAGCCCAAGCCACTGCTGTGCTGCAATGAGACACTTCCAGACGAG 7552

RESULT 3
 ID ADN04304 standard; cDNA; 8561 BP.
 AC ADN04304;
 AC ADN04304;
 DT 01-JUL-2004 (first entry)
 DE Antipsoariatic cDNA sequence #352.
 DE de; gene; antipsoariatic; gene therapy; psoriasis; diagnosis.
 KW Homo sapiens.
 OS
 PN W02004028479-A2.
 PD 08-APR-2004.
 PF 25-SEP-2003; 2003WO-US030907.
 PR 25-SEP-2002; 2002US-0414006P.
 PA (GENTH) GENENTECH INC.
 PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT,
 PI Wu TD;
 DR WPI; 2004-305105/28.
 DR P-PSDB; ADN04305.
 PT New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 PT
 PS Claim 1; SEQ ID NO 698; 3069bp; English.
 XX
 XX
 CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polynucleotides of the invention.
 XX
 SO Sequence 8561 BP; 1862 A; 3033 C; 2525 G; 1141 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 8561
 Score: 13215.00 Matches: 2517
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-09-522-753-5 (1-2517) x ADN04304 (1-8561)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
DB 2 ATGTGGGCTCCACACAGCTTGGCACAGAGTGGAGGCGCACTGAGCCCGGCTACCG 61
QY 21 ProHisSerLeuSerTyrTrpValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 62 CCCACAGGCTTCTTACCCAGTGCAGATCGCCGACGACACAGCACTCGGGCTCTCTG 121
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 122 GAGTACCCAGCACCTCCGCGACTATGCTCCACCTGTGCGCGGAGCTTCATCTCCAG 181
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 182 CCCACGCGGCGGAGGCGCTCCGCTGCTGTGAGTTCCAGCCCGGAAATACAGCTCCAG 241
QY 81 GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGluLeuGlyLeuSerGluMet 100
DB 242 GAGCTCCACCTCGGCGCAGAGTCCCATCTATACCTGCGAGCTGGGAGTCCAGAGATG 301
QY 101 GluPheIleGluSerTyrArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 302 GAGTTCATTGAAAGCAAGCGCCTCGGCTAGAGCTGTCTGATCCCTGCTGGAGCCG 361
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrTyrAspArgSer 140
DB 362 TCACCCCTGCTGGCCACGCGGCGCAGCTGGGAGATCGAAGACTCACCAAGAGACGTAGC 421
QY 141 LeuThrGlyLeuLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 422 CTGACGGGCAACTGGAACCGGTGTCTCCCGCCAGCCCGCCACACTGACCTTGAGCTG 481
QY 161 GluLeuValProProArgLeuSerTyrGluIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGTGTGCGCCACAGGCTGTCCAAAGAGAGCTGATCCAAAGACGTGAGCGGAGAC 541
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerTyrLeuLeuLeuLeuGlnGlnIleLeu 200
DB 542 CGAGAGATCAACATGATGAGCAGACAGATCTTAAGCTGAAGAGAGAGCAGACAGCTG 601
QY 201 GluGluGlnAlaAlaLeuProProGluProGluLeuProValSerProProIleGlu 220
DB 602 GAGGAGAGGCTGCCAAGCGCCCGAGCTCGAAGACCGGTGTACCGCCGCCATCGAG 661
QY 221 SerTyrHisArgSerLeuValGlnIleIleTyrAspGluAsnArgTyrValGlnAla 240
DB 662 TCGAAGCACCGCAGCTGTGAGATCATCTACGACGAGAACCGGAGAGAGGCTGAGCT 721
QY 241 AlaHisArgIleLeuGluGlnIleLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 722 GCACATCGGATTCGGAAGGCTGGGCGCCGAGGTGAGCTGCGCTGTACAAACAGCCGC 781
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLeuValLeuAsnGlnAlaMetArgTyrLeuLeu 280
DB 782 TCCGACACCCCGCAGATCATGAAACATCAAAATTAACACAGCGCATGCGGAAAGACTA 841
QY 281 IleLeuTyrPheLeuArgArgAsnHisAlaArgTyrGluTyrLeuLeuLeuPheCysGln 300
DB 842 ATCTGTACTCTTAAGAGGAGGAATCACGTCCGAAACAATGAGAGAGAGTCTGCCAG 901
QY 301 ArgTyrAspGluLeuMetGluAlaLeuGluLeuValGluValGluIleGluLeuAsnAspPro 320
DB 902 CCTATGACACAGCTCATGAGGCGCTTGGAAGAAAAAGTGGAGGCGATCCAAAAACACCG 961
QY 321 ArgArgArgAlaLeuGluSerTyrValArgGluTyrTyrGluLeuGlnPheProGluIle 340
DB 962 CCGCGGCGGCGCAAGAGAGCAAGGTGCGCGAGTACTAGAAAAAGCAGTTCCCTGAGATC 1021
QY 341 ArgTyrGlnArgGluLeuGlnGlnIleArgMetGlnSerArgValGluIleArgGlySerGly 360
DB 1022 CCGAAGCACCGGAGCTGAGAGCGCATGACAGAGGAGGTGGCCAGCGGCGGCACTGGG 1081

QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1082 CTGTCCATGTGGCGCGCCGACGACGACAGAGGTGTAGAGATCATGATGCTCTTCA 1141
QY 381 GluGlnGlnAsnLeuGluLeuGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1142 GAGCAGAGAACCTTGGAAGAGATGCGCAGCTGGCCGATATCCCGCCATGCTGTAC 1201
QY 401 AspAlaAspGlnGlnArgIleLeuPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1202 GACGCTGACACGACAGCGATCAAGTTTCAATCAACATGAAGAGGGCTTATGGCCGACCCATG 1261
QY 421 LysValTyrTyrAspArgGlnValMetAsnMetTrpSerGluGlnGluLeuGluThrPhe 440
DB 1262 AAGGTGTACAAAGACCGCAGGTCATGAACATGTGAGTGTAGACAGAGAGAGACCTTC 1321
QY 441 ArgGluLeuPheMetGlnHisProLeuAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1322 CCGAGAAAGTTCAATCACATCCCAAGAACTTTGGCTGATGCAATTCCTTGAGAGG 1381
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrTyrLysAsnGluAsnTyrLys 480
DB 1382 AAGACATGCTGAGATGGCTGCTCTATTACTACCTGACTGACAAAGAAATGAGAACTATAAG 1441
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1442 AGCTGTGTAGACGCGAGCTATGCGCGCGCCGCAAGACCGACCAACACAGCAGCAG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys 520
DB 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
QY 521 AspGluLysGluLysGluLysGluValGluLysGluGluGluGluGluGluGluGlu 540
DB 1562 GATGAG 1621
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
DB 1622 GACAAAGAAAGCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1682 AAGAGGCTGTGTGCTCCAAAGCGCCAAAGCTGCCAAAGCAGCAGAGAGAGAGAGAG 1741
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1742 CGCATCACCCGCTCAATGCTTAATGAGGCCAAACAGAGAGGCGATACCCCCAGCAG 1801
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620
DB 1802 AGCGCCGAGCTGGCTCCATGAGCTGAATGAGATTTCTCGCTGACAGAGAGAGAGAG 1861
QY 621 GluThrAlaLeuLysGlyLeuLeuGlnIleHisGlyAsnMetSerAlaIleAlaArgMet 640
DB 1862 GAAACAGCCAAAGAAAGGTCTCTGGAACACGCGCCCAATGTCGCAATCCCCGAGATG 1921
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
DB 1922 GTGGGCTCCAGACCTGTGTGCGAGGTAAAGAACTTCACTTCAATCAAGAAAGAGCAG 1981
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680
DB 1982 AACCTCGATGAGATCTTGACAGCAGCAACAGTGAAGATGGAAGAGAGAGAGAGCGCG 2041
QY 681 ArgTyrLysLysLysAlaProAlaAlaAspArgGluGluAlaPheProProValVal 700
DB 2042 AGAAGAGAGAGAGAGAGCGCGCGCGCGCGCAGCAGAGAGGCTGCATTCGCGCCGTGGTG 2101
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
DB 2102 GAGGATGAGAGATGAGAGCGCTCGGCGGTGAGCGGAATGAGAGAGAGATGTGTGAGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740

Db 2162 GGTGAAGCCCTTACATGCTCTGGGAATGAGTGCAGAGGGGATCAGTGGCCAGCC 2221
Qy ThrValAsnAsnSerSerAspThrGluSerIleProSerProHistiThrGluAlaIleAsp 760
Db 2222 ACTGTCAACAACAGCTCAGACCCGAGAGCATCCCTCTCTCTCACTGAGCGGCCAAG 2281
Qy AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2282 GACACAGGGCAGATGGGCCCAAGCCCCCAGCCACTCTGGGGCGCGAGCGGCCACCCCA 2341
Qy 781 GlyProProThrProProAlaGlyGlyThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2342 GGGCCACCCCAACCCCAACGAGAGACATCCCGGGCCCCCATTTGAGCCACCCGGCTCT 2401
Qy GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2402 GAAGCCACCGGAGCCCTTACGCCCCCAACGACACCCCATGCTCTGCACTCTCTCT 2461
Qy ValValProLysGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 840
Db 2462 GTGTCTCCCAAG 2521
Qy 841 GluGluGluGlyProProAlaAlaGlyGlyLeuAlaIleProThrGlyValGlyGly 860
Db 2522 GAGGACAG 2581
Qy ProValLysSerGluGlySerThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 880
Db 2582 CCGGTCAAG 2641
Qy 881 GluAlaAlaGlyAlaThrAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 900
Db 2642 GAGGCGCGTGAAGCCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2701
Qy 901 GlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThr 920
Db 2702 GGGAGGCGCACCACTGCAAGAGCTGGGGGCCCCCAGAGAGAGAGAGAGAGAGAGAG 2761
Qy 921 CysSerAlaAspGluValAspGluAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 940
Db 2762 TGGACGTGACAG 2821
Qy 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro 960
Db 2822 AGGCCCAAGCTCTCAACCCGAGCTGGCGAGACCCCGGGCCATCTCAACCCAGAGCCA 2881
Qy 961 LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLys 980
Db 2882 CTGGACCTGAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2941
Qy 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Db 2942 GTCCATGAGCCCGCCGGAG 3001
Qy 1001 ProProGlnAsnLeuGlnProGlnSerAspAlaProGlnGlnProGlnLysSerProArg 1020
Db 3002 CCACCGCAAAACCTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
Db 3062 GGCAG 3121
Qy 1041 LysLeuProGlyLysAspProCysTyrThrSerGlyLeuProPheProAlaProProArg 1060
Db 3122 AGCTGCTGGGAG 3181
Qy 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro 1080
Db 3182 GAGGTATCATAGAGCTCCCGGATGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3241
Qy 1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100

Db 3242 GGTCAACCACTGCGCTGGGCTTCATGACATGCCCGGCCGCTCTCGCGGCCACCC 3301
Qy ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg 1120
Db 3302 ACCATCTCAACCGGCTCCCTCTCATCTCTGTCCAGAGAGAGAGAGAGAGAGAGAGAG 3361
Qy 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis 1140
Db 3362 CAATAGGTGCATCTCCCAAGAGATGTGGTCCAGCTCCAGTCCCGGTCTCAGAGCAT 3421
Qy 1141 AlaLysAlaProValGlyProValThrMetGlyLeuProLeuPheMetAspProLys 1160
Db 3422 GCCAAGGCGCGGTGGGCTCTGTACCATGGGGCTGCCCTGCCATGAGACCCCAAAAG 3481
Qy 1161 LeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyAlaGlyPro 1180
Db 3482 CTGGACCTTCAAGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3541
Qy 1181 ProGlnSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
Db 3542 CCGGAGAGCTGGGGGGTCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3601
Qy 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
Db 3602 GGCTCAGTTCCGGGCGGAGAGATCACCAAGGATTCACAGACAGCGGTGCTCGGAC 3661
Qy 1221 SerAlaIleThrTyrThrGlySerIleThrHisGlyThrProAlaAspValLeuTyrLys 1240
Db 3662 AGCGCATCATCAACCGGCTCATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3721
Qy 1241 GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgIleArgGlu 1260
Db 3722 GGCACATCAACAGATCATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3781
Qy 1261 AspSerLeuProLysGlyHisValIleTyrGlyGlyLysValLeuLysSerTyr 1280
Db 3782 GACAGCTTCCCAAGGGCCAGCATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3841
Qy 1281 GluGlyGlyMetSerValThrGlnCysSerIleGluAspGlyValArgSerSerGlyPro 1300
Db 3842 GAGGAGGAGATGTCTGTGACCATGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3901
Qy 1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluIleArgValGlyArg 1320
Db 3902 CCCCATGAG 3961
Qy 1321 AlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db 3962 GCCATCTCTCAGAGAGATCGAAGGCTCATGGGCGGTGCCATCCCGCGGAGAGACAC 4021
Qy 1341 SerProHisIleLeuLysGlnGlnHisIleIleArgGlySerIleThrGlnGlyIlePro 1360
Db 4022 AGCCCCACACCTCAAAAG 4081
Qy 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380
Db 4082 CGGTCTTATGAG 4141
Qy 1381 GluGlyThrProProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 4142 GAGGAG 4201
Qy 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlyGlyLeuValAlaThrValLysGlu 1420
Db 4202 GCGCTGGAG 4261
Qy 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db 4262 GCGGGCGGCTTCATCATGATCCCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4321
Qy 1441 LeuAlaProArgProLeuLysGlyGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
Db 4322 CTGGCCCGCGGCGCTCAAG 4381

QY	1821	ValGluHhSaIaProlIetPhArgProGlyThrcIuInIserSerGlySer	1840
Db	5462	GtGAGAGACCGACCCATCTTGAGACCTGGTACAGAGACGACGGCAGGGCAGC	5521
QY	1841	SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis	1860
Db	5522	AGCGGCGGGGTGGGGGCGAGACGACGGCCGCCCTCCACTCCCATGCCACAGAC	5581
QY	1861	SerProIleSerProArgThrcInAspAlaLeuGlnGlnArgProSerValLeuHisIleAsn	1880
Db	5582	TCGCCCATCTCCCTCGGAGCCAGATGCTCCACAGAGAGACCAGTGTCTTCACAAAC	5641
QY	1881	ThrcGlyMetIysGlyIleIleThrcAlaValGlnProSerIysProThrValLeuArgSer	1900
Db	5642	ACAGGCATGAAGGGATCATCACCCTGTGGAGCCACAGAGCCACAGGTCTTAGGGTCC	5701
QY	1901	ThisSerThrsSerSerProValArgProAlaAlaThrcPheProProAlaThrcHisCysPro	1920
Db	5702	ACCTTCACCTCTCATCCCGTTCCGCCACGTCCCAATTCCCACTGCCACCCACTGCCCA	5761
QY	1921	LeuGlyGlyThrcLeuAspGlyValTyrcProThrcLeuMetGlnProValLeuLeuProIys	1940
Db	5762	CTGGGCGGCACCTCGATGGGGGTCTAACCTCACTCAATGAGCCCGTCTTGCTCCCAAG	5821
QY	1941	GluAlaProlArgValAlaArgProGlnArgProArgAlaAspThrcGlyHisAlaPheLeu	1960
Db	5822	GAGCGCCCCCGGGTCCGCCCGGACAGAGGGCCCCGAGACACCCGCAATGCTTCTCTC	5881
QY	1961	AlaIysPProProAlaArgSerGlyLeuGlnProAlaSerSerProSerTyGlySerGlu	1980
Db	5882	GCCAAAGCCCCAGCCCGCTCCGGGCTGAGGCCCGCTCTCTCCCAAGCAGGGCTCGGAG	5941
QY	1981	ProArgProLeuValProProValSerGlyHisAlaThrcIleAlaArgThrcProAlaIys	2000
Db	5942	CCCGGCGCCCTAGTGGCTCTGTCTGTGGCCAGCCACATCGCCGACCCCTGCGAAG	6001
QY	2001	AsnLeuAlaProHisHisAlaSerProAspProProAlaPProProAlaSerHisSerAsp	2020
Db	6002	AACCTCGACCTCAACCAAGCCAGCCCGGACCCGGCGGCCACCTGTCCGACTTCGGAC	6061
QY	2021	ProHisArgGlyIysThrcGlnSerIysPProPheSerIleGlnGlnLeuGlnLeuArgSer	2040
Db	6062	CCGACCCGGGAAAAGACTCAAAAGTAAACCTTTTCATCCAGAACTGGAACTCCGTTCT	6121
QY	2041	LeuGlyTyHisGlySerSerTySerProGlnGlyValGlnProValSerProValSer	2060
Db	6122	CTGGGTTCACAGCGACGACCTACAGCCCGGAAGGGGTGAGCCCGTCAGCCCTGTGAGC	6181
QY	2061	SerProSerLeuThrcHisAspIysGlyLeuProIysHisLeuGlnGlnLeuAspIysSer	2080
Db	6182	TCACCCCAATCTGACCCACAGACAAAGGGGTCTCCCAAGCACTCGGAAGAACTGACAAAGGC	6241
QY	2081	HisLeuGlnGlyGlnLeuArgProIysGlnProGlyProValIysLeuGlyGlyGlnAla	2100
Db	6242	CACCTGGAGGGAGACTCGGGCCCAAGGACGACAGGCCCGCGTGAAGCTTGGCGGGAGGCC	6301
QY	2101	AlaHisLeuProHisLeuArgProLeuArgProIysSerGlnProSerSerSerProLeuLeu	2120
Db	6302	GCCCACTCCCAACCTCGCGCGCGCTGCTGAAGCCAGCCCTGTCCAGCGCCGCTGTC	6361
QY	2121	GlnThrcAlaProGlyValIysGlyHisGlnArgValValThrcLeuAlaGlnHisIleSer	2140
Db	6362	CAGACCGGCCCAAGGGGTCAAAAGTCAACAGAGGGGTGTCACCTTGCCACACATCACTAGT	6421
QY	2141	GluValIleThrcGlnAspTyrcThrcArgHisIleProGlnGlnIleSerAlaProLeuPro	2160
Db	6422	GAGGTCAATCAACAGAACTTACACCCGGAGCAACCAACGACAGCTTAGCGACCCCTTGCCC	6481
QY	2161	AlaProLeuTyrcSerPheProGlyValaSerCysProValIleAspLeuArgArgProPro	2180
Db	6482	GCCCCCTCTACTCTTCTCCCTGGGGCCAGCTGCCCCGTCCTGAGACTCCGCGGCCACCC	6541
QY	2181	SerAspLeuTyrcLeuProProProAspHisGlyValaProAlaArgGlySerProHisSer	2200

```

Db 6542 AGTGAACCTTACCTCCGCGCCCGGACCATGTGCTCCCGGCTCCCGCCACAGC 6601
Qy 2201 GUGUGLYGYSASERPROGLUPROANLYRTHSERVALLEUGLYGLYGLUASP 2220
Db 6602 GAAAGGGGACAGAGCTCTCCAGACCAACAGCGTGGCTTGGGGTGGTGGAGAC 6661
Qy 2221 GUYLIEGLUPROVALSERPROGLUGLYMETRTHGLUPROGLYHISERASERALA 2240
Db 6662 GGRATTACACTGTGTCCCAACGGAGGCACTGAACGAGCCAGGCACTCCCGAGTCT 6721
Qy 2241 VALTYRPROLEULENTYRARGAPGLYGLUINLTHGLUPROSERARMETGLYSERLYS 2260
Db 6722 GTGTACCCGCTGTGTACCGGGATGGGGAAACAGACGAGCCACAGATGGCTCCAG 6781
Qy 2261 SERPROGLYANTRHSERGINPROPROALA PHEPHESELYLEUETHRGLUSERANSE 2280
Db 6782 TCTTCAGGCAACACCAACCCAGCCGCTTCTTACAGCAAGCTGACCGAGCACTCC 6841
Qy 2281 ALAMEVALYSESELYSLYSGINGLULAEANLYSLYSEUANTHRHISANARASN 2300
Db 6842 GCCATGTCTAAGTCCAGAGCAAGATCAACAGAAAGCTGAACACCCACACCGGAA 6901
Qy 2301 GLUPROGLUTYRANLIESERGINPROGLYTHRGULIEPHEANMETPROALALIETHR 2320
Db 6902 GAGCTGAATATACATATACAGCCAGCTGGGACGAGATCTTCAATATGCGCATCACC 6961
Qy 2321 GUYTHRGLYLEUWETHTYRARGSERGLALALINGLULHISALASERTHRANMET 2340
Db 6962 GGAACAGGCTTATGACCTTAGAAGCAGCGGTGACGAGAACTGCACACCAACATG 7021
Qy 2341 GUYLEUGLYUALLEILEARGLYSLALEUWECGLYRSTYRABPGINTPTGLUGLUSER 2360
Db 7022 GGGCTGAGGCGCATATTAAGAAAGCACTCATGGGTAAATATGACCTGGAGAACTCC 7081
Qy 2361 PROPROLEUSERALAENALAPHEANPROLEUANALASERLASERLEUPROALALA 2380
Db 7082 CCGCGCTCAGCGCCAAATGCTTTAACTTGATGCTGATGCTGACCTGCGCGCTGCT 7141
Qy 2381 METPROLETHRRLAALAAPGLYARGSERASPHISHTREUTHRSERPROGLYGLYGLY 2400
Db 7142 ATGCCCATTAACCGCTGTGACGAGGAGTGAACCACTCACTTGCAGAGTGGCGGC 7201
Qy 2401 GUYLYGALALYVALSERGYATRGPROSERSERARGLYVALYSESERPROALAPROGLY 2420
Db 7202 GGGAAAGGCCAAGTCTCTGGCAGACCCAGACCCGAAAGCCAAAGTCCCGGCGCGGC 7261
Qy 2421 LEUALISERGLYASPARPROPROSERVALSERSERVALHISERGLUGLYASPCYASN 2440
Db 7262 CTGGCATCTGGGAGCGGCAACCTGTCTCTCTCACTGCACTCGAGGAGACTGCAC 7321
Qy 2441 ARGARGTHRPROLEUTHRANARGVALTPGLUASPARPROSERSERALAGLYSERTHR 2460
Db 7322 CGCGGACCGCGTCCACCAACCGCTGTGGAGAGACAGCGCTCTCCGAGGTTCCACG 7381
Qy 2461 PROPHAPROTYRANPPROLEULIEMETARGLEUGLALAGLYVALMETALASERPRO 2480
Db 7382 CCATTCCCTCAACCCCTGATCATCGCGCTGACGGCGGTGTCTGCTTCCCAACC 7441
Qy 2481 PROPROGLYLEUBROALAGLYSERGLYPROLEUALAGLYPROHISALATPASPGLU 2500
Db 7442 CCACCGGGCTCCCGCGGGCAGCGGGCCCTCGCTGGGCCCAACCGCTGGGAGAG 7501
Qy 2501 GLUPROLYPROLEULENCYSERGINTYRGLUTHLEUSERASERGLU 2517
Db 7502 GAGCCCAAGCACTGCTGTGCTGACAGTACGAGCACTTCCGACAGCGAG 7552

```

```

DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF338 polynucleotide sequence SEQ ID NO:675.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatocytic;
KW vulnary; antipsoptic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antidiabetic;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN M0200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US006621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
DR P-PSDB; AAB40574.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX
PS Claim 5; Page 779-784; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatocytic; vulnary;
CC antipsoptic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antineumatic; antibacterial;
CC antiviral; antifungal; antineumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 8564 BP; 1859 A; 3034 C; 2532 G; 1139 T; 0 U; 0 Other;

```

Alignment Scores:

Pred. No.:	0	Length:	8564
Score:	13178.50	Matches:	2512
Percent Similarity:	99.80%	Conservative:	1
Best Local Similarity:	99.76%	Mismatches:	4

Query Match: 99.72% Indels: 1
DB: 3 Gaps: 1
US-09-522-753-5 (1-2517) x AAC74783 (1-8564)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
DB 1 ATGTCCGGGGCTCCACACAGCTTGTGGCAGACCTGAGGGCCACTAGAGCCCCCTCCGCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 61 CCCACAGCCCTTCTTACCCAGTGCAGATCGCCGAGCGCACAGAGCTGGGGCTCTCG 120
QY 41 GlnTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACAGACACACTCCCGGCACTATGCTCCACCTGTGCGCGGGCTCATCATCCAG 180
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGluValangluatgserGln 80
DB 181 CCCAGCGCGCGAGCGCTCTCCCTGCTGTGAGTTCAGGCTCCGGGAATGAACGGTCCAG 240
QY 81 GlnLeuHisLeuAspArgProGlnSerHisSerTyrLeuProGlnLeuGlyLysSerGlnMet 100
DB 241 GAGCTCACCTCGGGCGACAGTCCACTACCTGCTCCAGCTGGGGAAATCAGAGATG 300
QY 101 GlnPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAGAAGAGCGCCCTCGGCTTAGAGCTGCTGCTGACCCCTGCTGCGAGCG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 361 TCACCCCTCTGCGCCACGGGGCCAGCCTGGGATCTGAAGACCTCAACAAAGACCGTAGC 420
QY 141 LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGluLeu 160
DB 421 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCAGCCGCCGCACTGACCTTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerLysGlnGluLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGTCGCCGACGGCTGTCCAAAGAGAGACTATCCAGAACATGAGACCGCGTGGAG 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerLysLeuValLysValLysGlnGlnLeu 200
DB 541 CAGAGATATCACCTGTAGAGCAGCAGATCTTAACCTGAAGAAAGCAGCAACGACTG 600
QY 201 GlnGlnGlnAlaAlaLysProArgProGlnProGlnLysProValSerProProIleGln 220
DB 601 GAGGAGAGAGCTGCCAAGCCGCCGAGCCTGAGAACCCGCTGTACCCGCCCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 661 TCGAAGACACCGACGCTGTGGCAGATCATTTACGAGAGAACCGGAAGAGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnValGlnLeuProLeuTyrAsnGlnPro 260
DB 721 GCACATCGGATCTCGAAGAGCTCGGGGCCCCAGAGTGAAGTCCGCTGTACAACAGCCC 780
QY 261 SerAspThrArgGlnTyrHisSerLysAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TCCGACACCCGGGAGATCATGAGAACATCAAAATAAACAGGCGATGCGAAGAAAGCTA 840
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
DB 841 ATCTTCTACTTCAAGAGAGAGATCAAGCTCGAAGAACATGGAGAGCAAGATTCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysLysValGlnLysIleGlnLysAsnPro 320
DB 901 CGCTATGACACAGCTCATGAGGCGCTGGAAAAAAGTGAGAGCGCATTCGAGAACACCCC 960
QY 321 ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGluIle 340
DB 961 CGCGCGCGGGCCCAAGAGAGCAAGGTGCGGAGTACTACGAAAAAGCATTTCCCTGAGATC 1020

QY 341 ArgLysGlnArgGlnLeuGlnGlnLysMetGlnSerArgValGlyGlnArgGlySerGly 360
DB 1021 CCCAAGCAGCGGAGCTGCAGGAGCCCATGCAAGGCGGGTGGCCAGCGGGGCACTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTGGCGCCCGCAGCGAGCAGAGGTCTCAGAGATCATGAGCTCTCTCA 1140
QY 381 GlnGlnGlnLysLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1141 GAGCAGAGAACTTGAAGAGAGAGATGCCCAAGTGGCCGTGATCCGCCCATGCTGAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1201 GAGCTGACACAGCGCATCAAGTTCATCAATGAACGGGCTTATGGCCGACCCCAAG 1260
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnLysGlnThrPhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCATGAACATGTGAGTGAAGAGAGAGAGACCTTC 1320
QY 441 ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CGGAGAACTTCAATGAGCATCCCAAGAACTTGGCTGATGCAATCATTCCTGAGAGAG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnLysAntLys 480
DB 1381 AAGACAGTGGCTGAGTGCCTCTATTACTGACTGACTGAAGAAAGATGAAGACTTAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSer---GlnGlnGlnGlnGln 499
DB 1441 ACCCTGTGAGACGAGCTATGCGCGCCGGCAGAGCGACAGAGCAACACAGCAG 1500
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGln 519
DB 1501 CAGCAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
QY 520 LysAspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysProGluValGln 539
DB 1561 AAAAGATGAAGAGAGAAAGAAAGAGCGCGGAGAAAGAGAGAGAGAGAGCGGAGTGGAG 1620
QY 540 AsnAspLysGlnAspLeuLeuLysGlnLysThrAspAspThrSerGlyGlnAspAsnAsp 559
DB 1621 AACGACAAAGAAAGACTCTCTCAAGAGAGAGAGAGAGACCTCAGGGAGAGCAACGAG 1680
QY 560 GlnLysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnLysArgLys 579
DB 1681 GAGAGAGAGCTGTGCTCCAAAGGCCCAAACTGCCAAACAGCAGGAGAGAGACCCAA 1740
QY 580 GlyArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnGlnAlaIleThrProGln 599
DB 1741 GCGCGCATCACCCGCTCAATGGCTAATGAGGCCAACACAGAGAGAGGCAATCACCCCCAG 1800
QY 600 GlnSerAlaGlnLeuAlaSerMetGlnLeuAsnGlnLysSerArgTyrThrGlnGlnGln 619
DB 1801 CAGAGCGCGAGACTGCTCTCAATGAGCTGAATGAGATTCCTGCTGACAGAGAGAGAA 1860
QY 620 MetGlnThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArg 639
DB 1861 ATGGAACACGCCAAGAAAGTCTCTTGAAACACGGCCGAACTGCTGCGCATCGCCCG 1920
QY 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
DB 1921 ATGGTGGCTCCAAAGCTGTGTGCGAGTGAAGAACTTCACTTCAATCAAGAGAGAG 1980
QY 660 GlnAsnLeuAspGlnIleLeuGlnGlnHisLysLysLeuLysMetGlnLysGlnLysArg 679
DB 1981 CAGAACTTCGATGAGATCTTTCAGAGCAGCAAGAGCTGAATGAGAGAGAGAGAGAG 2040
QY 680 ArgArgArgLysLysLysAlaProAlaAlaAlaSerGlnGlnAlaAlaPheProProVal 699
DB 2041 CGAGAGAGAGAGAGAGAGAGCGCGGGGCGCGAGCGAGAGAGAGCTGATTCGCCCGCTG 2100
QY 700 ValGlnAspGlnGlnMetGlnLysSerGlyValSerGlyAsnGlnGlnGlnMetValGln 719

Dp	2101	GTGGAGGATGAGAGATGAGCGCTGCGGCGGAGCAAAATGAGAGAGATGGTGGAG	2160
Qy	720	GIuaIaGIuaIaLeuHiEaIaSerGIyAenGIuVaIProArGIyGIuCySerSerGIyPro	739
Dp	2161	GAGCGTGAAGCCTTACATGCTCTGTGGAAATGAGGTGCCAGAGGGGAATGCAATGGCCCA	2220
Qy	740	AlaThrValaAenAenSerSerAapThrGIuSerIleProSerProHiEThrGIuAlaA	759
Dp	2221	GCCACTGTCAACMAcAGCTCAGACACCGAGAGCATCCCTCTCTCCACACGGAGGCCGCC	2280
Qy	760	LysAapThrGIyGIuAenGIyProLyProProAlaThrLeuGIyAlaAapGIyProPro	779
Dp	2281	AAGGACACAGGCGACAAATGGGCCCCAAGCCCCCAGCCACTGGGGCGGAGCGGCGACCC	2340
Qy	780	ProGIyProProThrProProArGIrThrSerArGIaProIleGIuProThrProAla	799
Dp	2341	CCAGGCCCAACCCACCCACACAGAGAGACATCCCGGGCCCCCATTTAGACCCACCCGGCC	2400
Qy	800	SerGIuaIaThrGIyAlaProThrProProAlaProProSerProSerAlaProPro	819
Dp	2401	TCTGAAGGCACCGGAGCCCTTACGCCCCACACGACCCCATGCTCTGTGACCTCTCT	2460
Qy	820	ProValValaProLySGIuGIuLySGIuGIuTrAlaAlaAlaProProValaGIuGIu	839
Dp	2461	CCTGTGGTCCCCCAGAGAGAGAGAGAGAGACCGCAGCAGCCGCCCATGTGAGAGAG	2520
Qy	840	GIyGIuGIuGIuGIuLyProProAlaAlaGIuGIuLeuAlaValaAapThrGIyLyValaGIu	859
Dp	2521	GGGAGGAGACAGAAACCCCCCGGCGCTAGAGAGCTGGGACATGTGACACAGGAAAGCCAG	2580
Qy	860	GIuProValaLySerSerGIuCyEThrGIuGIuAlaGIuGIuGIyProAlaLySGIyLyAap	879
Dp	2581	GAGCCCGTCAAGAGAGAGATGACCGGAGGAAGCCGAGGAGGGCGGCCAAGGCAAGGAC	2640
Qy	880	AlaGIuAlaAlaGIuAlaThrAlaGIuGIyAlaLeuLyValaGIuLyLySGIuGIyGIy	899
Dp	2641	GCGGAGGCGCGTGAAGCCACCGGCCAGGGGGCGCTCAAGGACAGAAAGAGAGGCGCGG	2700
Qy	900	SerGIyArGIaIaThrThrAlaLySerSerGIyAlaProGIuAapSerAapSerSerAla	919
Dp	2701	AGCGGACAGGCGCACCTGCCAAGAGCTCGGGCGCCCCCAGAGACGGACCTCCAGTCTCT	2760
Qy	920	ThrCySerAlaAapGIuValaAapGIuAlaGIuGIyGIyAapLyAapArGIuLeuSer	939
Dp	2761	ACCTGCAATGCAAGAGAGTGGATGAGGCCGAGGGCGCGCGCAAGAACCGGCTGTGTCC	2820
Qy	940	ProArGIProSerLeuLeuThrProThrGIyAapProArGIaAenAlaSerProGIuLy	959
Dp	2821	CCAAAGGCCCACTCTCCACCCCGCACTGGCAGACCCCGGGCCAAATGTCTCACCCCAAGAG	2880
Qy	960	ProLeuAapLySGIuGIuLeuLySGIuAlaAlaAlaIleProProIleGIuValaThr	979
Dp	2881	CCACTGGAACCTGAAGACACTGAAGCAGAGCGCGCTGCATTCCTCCCATCCAGGTCAACC	2940
Qy	980	LyValaHiSGIuProProArGIuAapAlaAlaProThrLyProAlaProProAlaPro	999
Dp	2941	AAAGTCCATGAAGCCCCCGGAGAGACCAAGCTCCCAACCAAGCAAGCTCCCCAGCCCCA	3000
Qy	1000	ProProProGIuAenLeuGIuProGIuSerAapAlaProGIuGIuProGIySerSerPro	1019
Dp	3001	CCGCGCACCGCAAAACCTGCACACCGGAGAGCAGCGCCCTCTCAGACAGCTGGCAGAGCCCC	3060
Qy	1020	ArgGIyLySerArGIySerProAlaProProAlaAapLySGIuAlaAlaGIuAla	1039
Dp	3061	CGGGGCAAGACAGAGAGCCCGGACACCCCGCCGCAAGAGAGGCTTGTGACAGCGAGGCC	3120
Qy	1040	GIuLyLyLeuProGIyAapProProCyEThrThrSerGIyLeuProPhaProValaProPro	1059
Dp	3121	CAGAAAGCTGCTGGGGAGACCCCTTGTGTGACTTCCGGCCGTGCCCTTCCCGTGGCCCCC	3180
Qy	1060	ArgGIuValaIleLyValaSerProHiAlaAapAapProSerAlaApeSerTyAlaPro	1079
Dp	3181	CGTAAAGGTATCAAGAGCCTCCCGGCATGCCCCGGAACCCCTTCAGCTTCTCTACGCTCCA	3240
Qy	1080	ProGIyHiAapProLeuProLeuGIyLeuHiAapThrAlaArGIuValaProArGIyPro	1099
Dp	3241	CCTGTGATCACCACTGAGCCCTGTGGGCTCTCATGACACTCGCGGCGCGTCTGCGCGCCA	3300
Qy	1100	ProThrIleSerAapProProProLeuIleSerSerAlaLyValaAapSerValaLeuGIu	1119
Dp	3301	CCCAACATCTTCAACCCCGCTCCCTCATCTCTCTCTCCAGAGCAACCCAGAGTCTTCAG	3360
Qy	1120	ArgGIuIleGIyAlaIleSerGIuGIuMeSerValaGIuLeuHiAapProTySerGIu	1139
Dp	3361	AGGCAAAATAGTGCATCTCCCAAGAAATGTGGTCCAGCTCCACGTCCTGACTTCAGAG	3420
Qy	1140	HiAlaLyAlaAapProValaGIyProValaThrMetGIyLeuProLeuProMetAapProLy	1159
Dp	3421	CATGCCAAAGGCCCGGTGGGCGCTGTACCATCATTGGGCTGGCCCTGCCATGAGACCCAAA	3480
Qy	1160	LyLeuAlaAapProPhaSerGIyValaLyGIuGIuGIuLeuSerProArGIyGIuAlaGIy	1179
Dp	3481	AAAGTGCACCTTCAACGAGATGGAACAGAGACAGCTGTCCCAACGGGCGCAGGCTGGG	3540
Qy	1180	ProProGIuSerLeuGIyValaProThrAlaGIuGIuAlaSerValaAapGIyThrAla	1199
Dp	3541	CCACCGAGAGCCTCGGGGGGTGCCACAGCCCAAGAGGCGTCCGTGTGAGAGGAGACACT	3600
Qy	1200	LeuGIySerValaProGIyGIySerIleThrLySGIyIleProSerThrArGIaProSer	1219
Dp	3601	CTGGGCTCAGTCCGGGCGGAAGCATCAACAAAGCATTTCCAGACACACGGGTCCCTCG	3660
Qy	1220	AapSerAlaIleThrTyArGIySerIleThrHiSGIyThrProAlaAapValaLeuTy	1239
Dp	3661	GACAGCCCATCATATCCCGGCTCATATCAACCAACGACGACCACTGACGTCTCTATAC	3720
Qy	1240	LySGIyThrIleThrArGIyIleGIyGIuAapSerProSerArGIuAapArGIyArGIy	1259
Dp	3721	AAAGGCAACATCAACAGAAATCATCGGAGAGACGCCGAGTGTGCTGTGACCGGGCGCG	3780
Qy	1260	GIuAapSerLeuProLySGIyHiAapAlaIleTyGIuGIyLyLyLySGIyHiAapLeuSer	1279
Dp	3781	GAGGACAGCCTGCCCAAGGCGCACGTATCTTACGAAGAGCAAGAGGCCACGTCTTCTCC	3840
Qy	1280	TyGIuGIyGIyMeSerValaThrGIuCySerLySGIuAapGIyArGIySerSerSerGIy	1299
Dp	3841	TATAGAGGTGACATGTCTGTGACCCAGTGTCTCCAAAGAGAGCGGCAGAAACAGCTCAGCA	3900
Qy	1300	ProProHiSGIuThrAlaAlaAapProLyAaGIyThrTyAapMeThrGIuGIyArGIyGIy	1319
Dp	3901	CCCCCCCATGAAGAGGGCGGCCCAAGCGCACTTATACATGATGAGGGCGCGGTGGGC	3960
Qy	1320	ArgAlaIleSerSerAlaSerIleGIuGIyLeuMeGIyTyArGIaAlaIleProProGIuArGIy	1339
Dp	3961	AGAGCCATCTCTCAGCCAGCATGGAAGTCTCATGGAGCCGTGCCATCCGCGGAGCGA	4020
Qy	1340	HiSerProHiAapLeuLySGIuGIuHiAapIleArGIySerIle	1359
Dp	4021	CACAGCCCCACACCACTCAAAAGAGACACACATCCGGCGGTCAATCACACAAAGGATC	4080
Qy	1360	ProArGIyTyValaGIuAlaGIuAapTyLeuArGIuArGIuAlaLyLeuLeuLy	1379
Dp	4081	CCTGTGGTCTTACGTGAGGACAGAGAGACTTACTGTGTGGAGGCGCAAGCTCTTAAAG	4140
Qy	1380	ArgGIuGIyThrProProProProProProProSerArGIyAapLeuThrGIuAlaTyLyEThr	1399
Dp	4141	CGGAGAGGACAGCTCGGCCCCACCGCCTTCACAGGACCTTGACCGAGGCTTACMAAGC	4200
Qy	1400	GIuAlaLeuGIyProLeuLyLeuLyProAlaHiSGIyLyLeuValaIaThrValaLy	1419
Dp	4201	CAGGCTCTGGGCCCCCTGAAGCTGAAGCCGGCCCATATAGGGCTGTGGGCGCACGAGTGAAG	4260
Qy	1420	GIuAlaGIyArGIySerIleHiSGIuIleProArGIuGIuLeuArGIyIleThrProGIuLeu	1439
Dp	4261	GAGCGGGCGGCTTCATTCATGAGATCCCGCGGAGAGGCTGGGCGCACGCGCCGAGACTG	4320

QY 1440 ProLeuAlaProArgProLeuylsGluGlySerIleThrGlnGlyThrProLeuLysTYR 1459
DB 4321 CCCCTGGCCCCGGCGGCTCAAGAGAGGCTTCATACGAGGACCCCGCTCAAGTAC 4380
QY 1460 AspThrGlyAlaSerThrThrGlySerIleLysIleAspValAlaSerLeuIleGlySer 1479
DB 4381 GAAACGGGGCGGCTCCACCACTGGCTCCAAAGACAGACGTAAGCTCCCTCATCGGCAGC 4440
QY 1480 ProGlyArgThrPheProProValIleProLeuAspValMetAlaAspAlaArgAlaLeu 1499
DB 4441 CCGGGCCGAGACGTTCACACCCCTGCACCCGCTGATGTATGACCGACGCGCGGCACTG 4500
QY 1500 GluArgAlaCysTyrGluGlnSerLeuLysSerArgProGlyThrAlaSerSerSergly 1519
DB 4501 GAAAGTGGCTGCTACAGAGAGACCTGAAGAGCCGGCCAGAGACCCGACGACCTCGGG 4560
QY 1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProAlaGlnSer 1539
DB 4561 GGCTCATTTGGCGGCGCGCGCGCTGATGTGCTGAAGCTGGGTAAAGCGCGGAGAGC 4620
QY 1540 ProLeuThrThyGlyAlaAspHisGlyAlaProPheAlaGlyIleLeuProArgGlySerPro 1559
DB 4621 CCCCTAACTATAGAGAACCAAGGGACACCTTTGGCGGCACTCCACGAGGTTGCGCC 4680
QY 1560 ValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerglyAla 1579
DB 4681 GTGACCAAGCGGGAGCCCAAGCGCGCTGCAGAGAGGAGCAGCTTCGTGCACAGACCA 4740
QY 1580 SerGlnAspArgLysLeuThrSerThrProArgGlnIleAlaLysSerProHisSerThr 1599
DB 4741 TCCCAAGACCGAAAGGTGAGCTGACGCTCTGGAATCGCAAGTCCCGCACACACC 4800
QY 1600 ValProGlnHisIleProHisProIleSerProTyrGlnHisLeuLeuArgGlyValSer 1619
DB 4801 GTGCCCGAGACCAACCAACCCCATCTGCGCTTAAGAGCACTGCTTGGGGCGTGAAGT 4860
QY 1620 GlyValAlaAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg 1639
DB 4861 GGGGTGACCTGTATCGACAGCCATCCCGCTGCGCTTCGACCCCACTTCATACCCCGC 4920
QY 1640 GlyIleProLeuAspAlaAlaAlaAlaIleTyrTyrLeuProArgHisIleLeuAlaProAlaPro 1659
DB 4921 GGATCTCTGTGACCAAGCGCGCTGCTACTACTCCCGACACCTGGCGCCCAACCCC 4980
QY 1660 ThrTyrProHisLeuTyrProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu 1679
DB 4981 ACCTACCCGACCTGTACCACTTACTCTCATCCGCGCTAACCCCAAGAGCGCGCTG 5040
QY 1680 GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerglnGlnMetHisIleAsnThr 1699
DB 5041 GAGAAACGGGACGATCATCATGACTACATCACTCCGACGAGATGACCAACAAACAG 5100
QY 1700 AlaThrAlaMetAlaGlnIleArgAlaAspMetLeuArgGlyLeuSerProArgGlySerSer 1719
DB 5101 GCCACCGCATAGCGCCAGAGAGCTGATATGCTGAAGGGCGCTTCGCGCGCGAGTCTCG 5160
QY 1720 LeuAlaLeuAsnTyrAlaIleGlyProArgGlyIleIleAspLeuSerglnValProHis 1739
DB 5161 CTGGGACTAACTACCTGCGGAGTCCCGAGAGCATATGACTGTCCCAAGTGCACAC 5220
QY 1740 LeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAla 1759
DB 5221 CTGGCTGTGCTGTGCGCCGACACAGGACCCGACCGACCGCATGAGCGCTTGGCC 5280
QY 1760 TyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerPro 1779
DB 5281 TACCTCCCAACCGCGCGCGCTTCAGAGCGCGCACAGCAGCTCCCACTCTCCCA 5340
QY 1780 GlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerglnArgGluArgAsp 1799
DB 5341 GGAGGTCCAAACACTTGAACAAACACCAACGACTCTCTGTCGAGGCGGAGGAGAC 5400

QY 1800 ArgAspArgGluArgAspArgAspArgGluArgLysSerIleLeuThrSerThrThr 1819
DB 5401 CCGGATCGAGACCGGAGCCGGATCGGAGCCGGGAAATCCATCTCTCACTGCACAG 5460
QY 1820 ThrValGlnHisAlaProIleThrArgProGlyThrGlnGlnSerSerglySerSergly 1839
DB 5461 ACGGTGAGACGACCACTCTGAGACTGGTACAGACAGACAGCCGAGACAGCCG 5520
QY 1840 SerSerglyGlyGlyGlySergSerSerArgProAlaSerHisSerHisAlaIleGln 1859
DB 5521 AGCAGCGGGGGGGTGGGGGAGACAGACCGCCCGCTCCCACTCCATGCCACGACAG 5580
QY 1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHis 1879
DB 5581 CACTGCGCATCTCCCTTCGAGACCCAGGATGCGCTTCACAGACAGACCCGATGTCTTAC 5640
QY 1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
DB 5641 AACACAGGATGAAGGTATCATCACGCTGTGAGGCCAGCACCGCCAGGTTCTGAGG 5700
QY 1900 SerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys 1919
DB 5701 TCCACTCCACCTCTCTCAACCGGTTGCGCGGCTGCACATTCACCTGCACACCTGC 5760
QY 1920 ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuPro 1939
DB 5761 CCACTGGGGGAGCACTCGATGGGTCTACCTTACCTCATGAGAGCCCTTGTGCTGCC 5820
QY 1940 LysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPhe 1959
DB 5821 AAGAGAGCCCCCGGGTCCCGGCGAGAGCGGCCCGGAGACAGACCGGCAATGCTTTC 5880
QY 1960 LeuAlaLysProProAlaArgSerglyLeuGlnProAlaSerSerProSerLysGlySer 1979
DB 5881 CTGGCCAAAGCCCCAGCCCGCTCCGGGCTGAGCGCGCTCTCCGCCAGAGGCTCG 5940
QY 1980 GluProArgProLeuValProProValSerglyHisAlaThrIleAlaArgThrProAla 1999
DB 5941 GAGCCCCGGCCCCCTGAGCTCTCTGTCTGTGACAGCCACCATGCCCCGACCCCTGG 6000
QY 2000 LysAsnLeuAlaProHisIleAlaSerProAspProProAlaProProAlaSerAlaSer 2019
DB 6001 AAGAACTCTGACCTTACACCAAGCCGAGCCGAGCCGCGGCACTCTCTGAGCTCG 6060
QY 2020 AspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGlnLeuArg 2039
DB 6061 GACCCGCAAGGGGAAAGCTCAAGTAAACCTTTTCATGACAGAACTGGAACCTCGT 6120
QY 2040 SerLeuGlyTyrHisGlySergSerTyrSerProGlnGlyValGluProValSerProVal 2059
DB 6121 TCTTGGGTTTACAGCGGAGAGCTTACAGCCCGGAAGGGGTGAGCCGCTCAGCCCTGTG 6180
QY 2060 SerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlnGluLeuAspLys 2079
DB 6181 ACCTACCCAGTCTGACCAACCAAGAGGGGCTCCCAAGACACTGAGAGCTCGACAAG 6240
QY 2080 SerHisLeuGlnGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGlu 2099
DB 6241 AGCCACCTGAGAGGGAGGTGGGCGCCAAAGCCAGCGCCCTGTAAGCTTGGCGGAGG 6300
QY 2100 AlaAlaHisLeuProHisLeuArgProLeuProGlnSerGlnProSerSerSerProLeu 2119
DB 6301 GCGGCCACCTCCACACTCGCGCGGCTGTGAGAGCAACCTCGTCCAGCCGCGTG 6360
QY 2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle 2139
DB 6361 CTTCAGACCGCCCGAGGGGTCAAGAGTCAACAGCGGTGTCACTTGGCCAGACATC 6420
QY 2140 SerGluValIleThrGlnAspTyrThrArgHisIleProGlnGlnLeuSerAlaProLeu 2159
DB 6421 AGTGAGGTATACACAGAGACTACACCGGACACCAACAGACACTGACCGCACCCCTG 6480
QY 2160 ProAlaProLeuTyrSerThrPheProGlyAlaSerCysProValLeuAspLeuArgArgPro 2179

```

Db      6481  CCGGCCCCCTACTCTTCCCTGGGCGACGCTGCCCTCCGACCTCCGCGCCCA 6540
Qy      2180  ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis 2199
Db      6541  CCGAGTACCTTACCTCCGCGCCCGGACCATGTGCCCCGGGCTGGCTCCCGCCAC 6600
Qy      2200  SerGluGlyGlyAspSerProGluProAsnLysSerValLeuGlyGlyGlu 2219
Db      6601  AGCGAAGGGGCAAGAGCTCTCCAGAGCAACAGAGCTGCTGGGTGGGTGGAG 6660
Qy      2220  AspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSer 2239
Db      6661  GACGGTATTAACTGTGTCCCGCCCGAGGGCATGACGGAGCGAGCGACCTCCCGAGT 6720
Qy      2240  AlaValThrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db      6721  GCTGTGATCCCGCTGCTGTACCGGGATGGGAGACAGACGGAGCCAGAGATGGGCTCC 6780
Qy      2260  LysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluSerAsn 2279
Db      6781  AAGTCTCAGGCAACACACAGCCCGCCAGCTTCTTACAGCAAGCTGACCGAGAGCAC 6840
Qy      2280  SerAlaMetValLysSerLysGlyGlnGluIleAsnLysLysLeuAsnThrHisAsnArg 2299
Db      6841  TCCGCCATGTGCAAGTCCAGAGCAAGATCAACAAGAGCTGACACCCACCAACCGG 6900
Qy      2300  AsnGluProGluTyrAsnLysSerGlnProGlyThrGluIlePheAsnMetProAlaIle 2319
Db      6901  AATGAGCTTAATATCAATATCAGACAGCTGGGAGCGAGATCTTCAATATGCCCCATC 6960
Qy      2320  ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
Db      6961  ACCGGAAACAGGCTTATGACTTATGAGACCAAGCGGTGACGAGAACTGCCAGACCAAC 7020
Qy      2340  MetGlyLeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
Db      7021  ATGGGGCTGAGGCGCATATTAAGAAAGCACTCATGTGTAATATGACCAAGTGGAGAG 7080
Qy      2360  SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLysLeuProAla 2379
Db      7081  TCCCGCGCTCAGCGCGCAATGCTTTTAACTCTGTAATGCCATGCCACCTGCCGCT 7140
Qy      2380  AlaMetProLethrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2399
Db      7141  GCTATGCCCATACCGCTGCTGACGAGCGAGTGAACACACTCACCCTGCGCAGGTGC 7200
Qy      2400  GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
Db      7201  GGGGGGAAGGCAAGGCTCTGGGAGACCCAGACGCCGAAAGCAAGTCCCGGCGCCG 7260
Qy      2420  GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
Db      7261  GGGCTGGCACTCTGGGAGACCGGCCACCTCTGTCTCCCTCACTGCGAGGGGAGACTGC 7320
Qy      2440  AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db      7321  AACCGCGGAGCGGCTCACCACACCGGTGTGGAGAGACGCTCTGTCGCGAGGTTCC 7380
Qy      2460  ThrProPheProTyrAsnProLeuLysMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
Db      7381  AGCGCAATCCCTCCACAAACCCCTGATCATGGGTGACGGGTGTCAAGGCTTCCCA 7440
Qy      2480  ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
Db      7441  CCCCACACCGGGCTCCCGCGGAGCGGCGCCCTCCCTGCTGCGCCACCGCTGGGAC 7500
Qy      2500  GluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db      7501  GAGGAGCCCAAGCACTGCTCTGCTGCAAGTACAGAGCACTCTCCGACAGGAG 7554

```

```

ID      ACN44283 standard; cDNA; 8667 BP.
XX
XX
AC      ACN44283;
XX
XX
DT      18-NOV-2004 (first entry)
XX
XX
DE      Human mRNA sequence hCT16428.
XX
XX
KW      Cytostatic; carcinoma; lymphoma; cancer; human; gene; db.
XX
XX
OS      Homo sapiens.
XX
XX
PN      WO2003073826-A2.
XX
XX
PD      12-SEP-2003.
XX
XX
PF      28-FEB-2003; 2003MO-US006235.
XX
XX
PR      01-MAR-2002; 2002US-00087192.
XX
XX
PA      (SAGR-) SAGRES DISCOVERY.
XX
XX
PI      Morris DW;
XX
XX
DR      WPI; 2003-328604/31.
XX
XX
PT      Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT      comprises a nucleotide sequence.
XX
XX
PS      Claim 1; SEQ ID NO 653; 0pp; English.
XX
XX
CC      The present invention relates to novel DNA and protein sequences which
CC      are associated with carcinomas. The sequences are useful for: (i) for
CC      screening drug candidates; (ii) for screening of bioactive agent capable
CC      of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC      a bioactive agent capable of modulating the activity of CAP; (iv) for
CC      evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC      carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC      carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC      (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC      determining Carcinoma Associated (CA) gene copy number. In addition, the
CC      CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC      carcinoma including lymphoma. The present sequence is one such CA gene
CC      sequence. Note: This patent is an equivalent to basic Patent
CC      US2002182586A1, for which no sequence data was published
XX
XX
SQ      Sequence 8667 BP; 1860 A; 3066 C; 2575 G; 1166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0
Score: 13145.00
Percent Similarity: 99.68%
Best Local Similarity: 99.64%
Query Match: 99.47%
DB: 11
Gaps: 0
Length: 8667
Matches: 4* 2509
Conservative: 1
Mismatch: 7
Indels: 2
Gaps: 0

US-09-522-753-5 (1-2517) x ACN44283 (1-8667)
Qy      1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro 20
Db      116 ATGTGGGATCCACACAGCCTGTGGACAGACTGAGGGGCACTAGAGCCCTACCCG 175
Qy      21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db      176 CCCACAGCCTTCTTCAACCAAGTGCAGATGCGCCGACGACAGCAAGAGCTGGGCTCTG 235
Qy      41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db      236 GAGTACACAGACACACTCCCGCACTATGCTCCACACTGTGCGCCGGCTCAATCAG 295
Qy      61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db      296 CCCACGCGGAGGCGCTCCCTGCTGTGATTCAGACCCGGGATGAACGGTCCAG 355

```


QY 81 GtUeUhiLeUaRgProGiuSerHisSerTyLeuProGluEugLyysSerGluMet 100
DB 356 GAGCTCACCTGGGGCCAGAGTCCCACTACCTGCCGAGCTGGGAAAGTCAGAGATG 415
QY 101 GtUphEiLeGiuSerTyRgRgProAglEugLiuELeuProAsPProLeuLeuAagPro 120
DB 416 GAGTTCAATTGAAAGCAAGGCCCTCGCTCAAGACTCTGCTGACCCCTGCTGCACCG 475
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThryAspAagSer 140
DB 476 TCAACCTCTGCTGGCCACGGGCCAGCTCTGGGAGTCTGAAGACTTCAACAGACCTTAC 535
QY 141 LeuThrGlyLyLeuEugLiuProValSerProProSerProProHisThryAspProGluLeu 160
DB 536 CTGACGGGCAAGCTGGAAACGGGTGTCTCCCGACGCCCGCACACTGACCTTGAGCTG 595
QY 161 GtUeUeValProProAglEugSerTyRgGluGluLeuTLeGluAspMetLeuAspValAsp 180
DB 596 GAGCTGTGTCGGCCACGGCTGTCCAAAGAGAGCTGATCCAGAACATGGACCCGCTGGAC 655
QY 181 ArgGluLeuThrMetValGluGlnGlnHisSerTyLeuLySylSylGlnGlnLeu 200
DB 656 CAGAGATCAACATGGTAGAGCAGCATCTTAAGCTGAAGAGAGAGACAGCAACAGCTG 715
QY 201 GtUgUgUaLalAlaLySProProGluProGluLySProValSerProProProLeGlu 220
DB 716 GAGAGAGAGGCTGCCAAGCGGCCGAGCTGAGAACCGGTGTCAACGCCGCCCATCGAG 775
QY 221 SerTyShiAspSerLeuValGlnHisSerTyAspGluAspGlySylSylAla 240
DB 776 TCGAAGCACCGCAGCTGTGAGATCATCTACAGAGAACCCGGAAGAGGCTGAAGCT 835
QY 241 AlaHisArgLeuEugLiuGlyLeuGlyProGluValGluEupProLeuTyRasGlnPro 260
DB 836 GACATCGAGATTCGGAAGGCTGGGGCCCGAGGTGAGCTGCGCTGTACACCAAGCCC 895
QY 261 SerAspThrArgGlnTyRhiSglUeUeNileysIleAsnGlnAlaMetArgLySylSylE 280
DB 896 TCCGACACCCGGCAGTATCATGAAACATCAAAATTAACAGGCGATGGCGAAGAAAGCTA 955
QY 281 IleLeuTyRheLySylAspArgAsnHisAlaArgLySglInTlySglInLySylSylGln 300
DB 956 ATCTTCTACTTCAAGAGGAGGAAATCAAGCTCGGAAACAATGGGAGAGAAATGTTCTCCAG 1015
QY 301 ArgTyRAspGluEueMetGluAlaEugLiuLySylValGluArgLileGluAsnAspPro 320
DB 1016 CGCTATGACCACTCATGAGGCTTGAGAGAAAGAGGTGAGCGCATCGAACAACCCC 1075
QY 321 ArgArgArgAlaLySglUeSerTySylValArgGluTyRtyGluLySglInPheProGluIle 340
DB 1076 CGCGCGCGGGCCAGAGAGAGCAAGTGGCGAGTACTACAGAAAGCATTCCTTGAGATC 1135
QY 341 ArgLySglInaRgGluEugGlnUaRgMetGlnSerArgValGlyGlnaRgLySergLy 360
DB 1136 CGCAACACACCGCAGCTGACGAGCCATGCAAGACAGGGTGGCCACGCGGCAAGTGGG 1195
QY 361 LeuSerMetSerLalaArgSerGluHisGluValSerGluTLeIleAspGlyLeuSer 380
DB 1196 CTGTTCATGTGGCCCGCCGACGAGCAAGAGTGTCAAGATCATGATGGCTCTTCA 1255
QY 381 GtUgUgUaLalAsnLeuGluLySglInMetArgGlnLeuAlaValIleProPomeTLeuTyR 400
DB 1256 GAGCAGAGAACTGGAGAAAGAGATGGCCGACACTGGCCGTGATCCCGCCCACTGCTTAC 1315
QY 401 AspAlaAspGlnGlnaRgLySylSylSylSylSylSylSylSylSylSylSylSylSylSyl 420
DB 1316 GAGCGTGAACAGAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1375
QY 421 LySValTyLySAspArgGlnValMetAspMetTySerGluGlnGluLySglInThryPhe 440
DB 1376 AAGGTGTACAAACCGCCAGTGCATGAAATGTGAGTGAAGAGAGAGAGAACCTTTC 1435

QY 441 ArgGluLySylPheMetGlnHisProLySylAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1436 CGGAGAAAGTTATGACAGATCCCAAGAACTTGGCTGTATGCATCATCTCTGGAGAG 1495
QY 461 LySThrValAlaGlySylValLeuTyRtyRtyRtyRtyRtyRtyRtyRtyRtyRtyRtyR 480
DB 1496 AAGACATGTGCTGATGTGCTCTCTTATCTTACTCTACATGAAGAAATGAAGACTTAA 1555
QY 481 SerLeuValArgAspSerTyRArgArgArgLySylSergGlnGlnGlnGlnGlnGln 500
DB 1556 AGCTGTGTAGACGAGATCATCGGCGCGGCAAGAGCCAGACCAACACAGCAGCAG 1615
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1616 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1675
QY 521 AspGluLySglUeSglUeLySglUeLySglUeLySglUeLySglUeLySglUeLySglUe 540
DB 1676 GATGAG 1735
QY 541 AspLySglUaSpLeuLeuLySglUeLySglUeLySglUeLySglUeLySglUeLySglUe 560
DB 1736 GACAGAGAGAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
QY 561 LySglUaLalAlaSerLySglTyArgLySylThryAspSerGlnGlyArgArgLySgl 580
DB 1796 AAGAGGCTGTGGCTCCAAAGCGCCGCAAAACTGTCCAAACAGCAGGAGAGAGC 1855
QY 581 ArgIleThrArgSerMetValAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1856 CGCATCACCGGCTCAATGGCTAATGAAGCCAAACAGGAGAGAGCCATCACCCCCAGCAG 1915
QY 601 SerAlaGluLeuAlaSerMetGluEuaGluSerSerArgTyRThryGluGluGluMet 620
DB 1916 AGCGCGAGACTGGCTCCATGAGCTGATGAGATGAGTTCGCTGAGACAGAAAGAAATG 1975
QY 621 GtUthrAlaLySylSylLeuEugLiuHisGlyArgAsnTyRserAlaIleAlaArgMet 640
DB 1976 GAAACAGCCAAAGAAAGTCTCTGGAACAAGCGCGCACTGGCTGCGCATCGCCGCAATG 2035
QY 641 ValGlySerTyRThryValSerGlnTySylSylAsnPheTyRAspThryLySylSylArgGln 660
DB 2036 GTGGGCTCCAAAGCTGTGTGCGAGTGTAAACTTCTTACTTCAACTACAAAGAAAGCAG 2095
QY 661 AsnLeuAspGluIleEugGlnGlnHisLySylSylSylSylSylSylSylSylSylSylSyl 680
DB 2096 AACCTCGATGAGATCTTGAAGCAGCAGCAAGCTGAAGATGAAGAGAGAGAAACGGCGG 2155
QY 681 ArgLySylSylSylAlaProAlaAlaAspGluGluAlaAlaPheProProVal 700
DB 2156 AGGAAGAAAGAAACCGCGCGCGCGCGCGCAGCGAGAGAGGCTGCATTCGCCCGCTG 2215
QY 701 GtUaSpGluGluMetGluAlaSerGlyValAspGlyAsnGluGluMetValGluGlu 720
DB 2216 GAGAGTAGAGAGATGGAGCGTGGCGGTGACCGGAAATAGAGAGAGAGATGGTGGAGAG 2275
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlyLucySergLyProAla 740
DB 2276 GCTGAAGCTTATATCTCTGTGGAAATGAGTGCCTCAAGAGGGAATGCATGGCCAGCC 2335
QY 741 ThrValAsnAspSerSerAspThryGluSerIleProSerProHisThryGluAlaAlaLyS 760
DB 2336 ACTGTCAACAAACAGCTCAACACCGAGACATCCCTCTCTCAACATGAGGCGCCAG 2395
QY 761 AspThrGlyGlnAsnGlyProLySylProAlaThrLeuGlyValAspGlyProProPro 780
DB 2396 GACACAGGCGCAGATGGGCCCAAGCCCGCAGCACCTTGCGGCCAGCGGAGCACCCCA 2455
QY 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
DB 2456 GGGCCACCCACCCACACCGAGAGAGATTCGGCC-CCCATGAGCCACCCCGGCTTC 2514
QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820

[illegible]

Db	3595	GCTGGACACCTTCACGGAGTGAACAGAGACAGCTGTCCCAACGGGGCCAGGCTGGGCC	3655
QY	1180	oProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgIYThrAlaLe	1200
Db	3655	ACCGGAGAGCCTGGGGGTGGCCACAGGCCACGAGAGGCGTCCGGTGTGAAGAGGACAGCTCT	3714
QY	1200	uGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAs	1222
Db	3715	GGGCTCACTTCGGGGCGGAGACATCACMAAGGCAATTCGACAGCACCGGGTCCCTCGGA	3774
QY	1220	pSerAlaIleThrTYrArgGlySerIleLeuHisGlyTYrProAlaAspValIleuTYrLys	1240
Db	3775	CAGCGCCATCAATACCGGGCTCCATATCCACGAGCAGCGCAGGTGACGTCTGTACAA	3833
QY	1240	sgLYThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgIYArgGly	1260
Db	3835	GGGACCATTCACAGGATCATCGGGAGGACAGCCCGAGTGGCTTTGACCGCGGCCGGGA	3894
QY	1260	uAspSerLeuProLysGlyHisValIleTYrGluGlyLysValGlyHisValIleuSerTY	1280
Db	3895	GGACAGCCTGGCCCAAGGGCCACGTCATCTAGAAAGGCAAGAGGGCCACGTCTGTCTTA	3954
QY	1280	rgLugLYGlyMetSerValIThrGlnGlySerLysGluuAspGlyArgSerSerArgLYPr	1300
Db	3955	TGAGGGGTGGCAAGTCTGTGACCGAGTGTCCAAAGAGACGGCAGAACACCTCAAGACC	4014
QY	1300	oProHisGluThrAlaAlaProLysArgTYrTYrAspMetMetGluGlyArgValGlyArg	1320
Db	4015	CCCCCATAGACGGCGCCGCCCAAGGACCACTTAGCATGTATGAGGGCCCGCGTGGCAG	4074
QY	1320	gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis	1340
Db	4075	AGCCATCTCCTCAGCCAGCATCGAAGGTCTCATGGCGCGTCCATCCGCGGAGCGACA	4134
QY	1340	sSerProHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIlePr	1360
Db	4135	CAGCCCCACCACTCAAGAGCAGCACCAATCCGGGGTCCATCAACAAAGGATCTCC	4194
QY	1360	oArgSerTYrArgIuAlaGlnGluuAspTYrLeuArgArgGlyuAlaLysLeuLeuLysAr	1380
Db	4195	TCCGTCTCTACGTGAGGACAGAGAGACTCTCGTCGGAGGCCAAGCTCTTAAACCG	4254
QY	1380	gGluGlyTYrProProProProProSerArgAspLeuHisArgIYArgTYrLysThrGly	1400
Db	4255	GGAAGGACCGCTCCGCCGCCACCGCCCTCAACGGACTGACGAGGCTTCAAGACGCA	4314
QY	1400	nAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLysValAlaThrValLysGly	1420
Db	4315	GGCCCTGGGGCCCCCTGAAGCTGAAGCGGGCCCAAGAGGGCTGTGTGGCACAGGTGAAGGA	4374
QY	1420	uAlaGlyArgSerIleHisGluIleProArgGluLeuArgHisThrProGluuLysPr	1440
Db	4375	GGCGGGCGGCTTCATCATGAGATCCCGCGGAGAGACTGTGGGACACGCGCGAGCTGCC	4434
QY	1440	oLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyTYrProLeuLysTYrAs	1460
Db	4435	CTTGCGCCCGGGCGGCTCAAGAGAGGGCTCATACCGAAGGCAACCGCGCTCAAGTACGA	4494
QY	1460	pThrGlyAlaSerThrThrGlySerLysValHisAspValArgSerLeuIleGlySerPr	1480
Db	4495	CACCGGCGGCTGCACCACTGGCTCCAAAAAGCACAGTACGCTCCCTCATCGGCACGCC	4554
QY	1480	ogLYArgTYrThrProProValHisProLysAspValMetAlaAspAlaArgAlaLugLY	1500
Db	4555	CGGCGGACGTTCCACCCCGTGCACCGCTGAGATGATAGCCACGCCCGGGACATGGA	4614
QY	1500	uArgAlaCysArgTYrGluGluSerLeuLysSerArgProGlyTYrAlaSerSerSerGlyGly	1520
Db	4615	ACGTCGCTGTCAAGAGAGAGCTGAAGAGCGGCGGACCGGCACGACGCTCGGGGGG	4674
QY	1520	ySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgLysPr	1540
Db	4675	CTTCATTTGGCGGCGGCGCCCGGTGCTATTTGGCTGGAGCTGGCAAGCCGCGGCAAGACC	4734

QY 1540 oLeuThrTYrgLUAspHISglYAlaProPheAlaGlYHISLeuProArgGlySerProVa 1560
DB 4735 CcTAACCTATGAGGACCAcGGGGcACcCTTgCCGGGcACcCTCCcAGAGGTTcGcCCGg 4794
QY 1560 lThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerIysAlaSe 1580
DB 4795 GACcACGcGGGAGcCCcACGcCGGcCTTGcAGAGGAGcAGcCTTcGTCCGcAGAGGcATC 4854
QY 1580 rGlnAserGlyLeuThrSerThrProArgGluIleAlaYSerProHISerThrVa 1600
DB 4855 CcAGGACcCGAAAGcTGAcGTCCGcAGcCGcCTGcAGATcCGcCAATcCCcCGcACGcACcCGT 4914
QY 1600 lProGluHISerHISerProHISerProIleSerProTYrgLuhISLeuLeuArgGlyValISergI 1620
DB 4915 GcCCcGAGcACcACcCCcACcCCcCATcCTCGcCTTcGAGcACcCTGcCTTcGGGGGcGTGAGTGG 4974
QY 1620 yValaLapLeuTYrArgSerHISerIleProLeuAlaPheAserProThSerIleProArgG 1640
DB 4975 GcTGAGcCTGTATcCGcAGcCCcATcCCcCTGcGGcCTTGcAGcCCcCATcCTCCcATcCCcCGG 5034
QY 1640 yIleProLeuAAspAlaAlaAlaIATYrTYrLeuProArgHISLeuAlaProAspProTh 1660
DB 5035 CATcCTcCTTGAGcAGcAGcCGcCTGcCTATcCTGcCCcCGcACcCTGcCCcCAACcCCcCAC 5094
QY 1660 rTYrProHISLeuTYrProProTYrLeuIleArgGlyTYrProAspThrAlaAlaLeuG 1680
DB 5095 CTACcCGcACcCTGTACcCCcCATcCTACcCTACcCGcGTACcCGcGTACcCGcAGcACcGGcCGcGTGG 5154
QY 1680 uAenAArgGlnThrIleIleAsnAAspTYrIleThrSerGlnGlnMetHISerAsnThrAl 1700
DB 5155 GAACcCGcAGcACcCATcATcCATcATcCATcATcCATcCTGcAGcCATcCATcCACcCAACcCGG 5214
QY 1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlySerSerIe 1720
DB 5215 CACcCGcCATcGGcCCcACcAGcACTATcGTcGAGcGGcCTTcCGcCCcCGcAGcTCTcGTcGT 5274
QY 1720 uAlaLeuAAspTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHISer 1740
DB 5275 GcGACTcCAACTcAGcCTGcGGGTcCCcCGcAGcCATcCATcGACTcGTCCcCAATcGcCACcCGT 5334
QY 1740 uProValleuValProProThrProGlyThrProAlaThrAlaMetAAspArgLeuAlaTy 1760
DB 5335 GcCTGTcGTCTcGTGcCCcCGcACcAGcAGcACcCCcAGcACcCGcCATcGAGcACcCGcCTTcGTcCTA 5394
QY 1760 rLeuProThrAlaProGlnProPheSerSerArgHISerSerSerProLeuSerProG 1780
DB 5395 CTcCCcCCcACcCGcCCcCCcAGcCTTcTAGcAGcCCcCGcACcAGcCTcCCcCACTcCTcCCcCAG 5454
QY 1780 yGlyProThrHISLeuThrIleAspProThrThrThrSerSerSerGluArgGluArgAspAr 1800
DB 5455 AGGTcCAACcACcACTTcGACAAACcCAACcCACcCGTcCTcGTcCGAGcGGGAGcGAGcACCG 5514
QY 1800 gAspAArgGluArgAspArgAspArgGluArgGlyIleuYSerIleLeuThrSerThrThrTh 1820
DB 5515 GcATTCAGAGcCGGAGcCGGATcGGAGcCGGAAAGTcCATcCTcCACTcCCcCGcACcCGcAG 5574
QY 1820 rValGluHISerAlaProIleTyrArgProGlyThrGlnInSerSerGlySerSerGlySe 1840
DB 5575 GcTGAGcAGcAGcACcCATcCTGcAGcACTcGTcACAGcAGcAGcAGcAGcAGcAGcAGcAGcAG 5634
QY 1840 rSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHISerHISerHISerAlaHISerGlnH 1860
DB 5635 CAGcCGcCGGGGcGTGGGGGcAGcAGcAGcCGcCGcCTcCCcCATcCTCCcCATcCCcCAGcAG 5694
QY 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValleuHISer 1880
DB 5695 CTcCGcCATcCTcCTcCTcCTcCGcACcCGAGATcCGcCTcCGcAGcAGcACcCAcGTGTcCTTcCA 5754
QY 1880 nThrGlyMetLeuGlyIleIleThrAlaValGluProSerSerProThrValleuArgSe 1900
DB 5755 CACAGcCATcGAAAGcGTATcATcACcCGcCTGTcGAGcCCcAGcACcCGcCCcCATcCGGTcCTGAGG 5814

QY 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHISerSer- 1920
DB 5815 CACTcCACcCTcCTcCTcCTcCTcCGcCTGcCCcGGcCTGcCATcATTcCCcACcCTGcCCcACcCGcCGCC 5874
QY 1920 oLeuGlyGlyThrIleuAAspGlyValTYrProThrIleuMetGluProValleuLeuProTy 1940
DB 5875 ACTGGcGGGACcCTcGTAGAGGgGTcTACcCTACcCTcCATcGAGACcCCcGTcCTTGTcGCTCCAA 5934
QY 1940 sGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHISerAlaPheIe 1960
DB 5935 GAGAGcCCcCCcGGGTcGGcCGcAGAGcGGcCCcGAGcAGcACcCGcGcCATcCTcCTcCT 5994
QY 1960 uAlaTyProProAlaArgSerGlyLeuGluProAlaSerSerProSerIysGlySerG 1980
DB 5995 CcCGcCAAGcCCcCGcAGcCGcCTcCGGcGTGAGcCGcCGcCTcCTcCCcCGcAGcAGGcCTcCGA 6054
QY 1980 uProArgProLeuValProProValSerGlyHISerAlaThrIleAlaArgThrProAlaTy 2000
DB 6055 GcCCcCGcCCcCTTAGcGTcCTcGTcCTcGTcGcCCcAGcCCcCATcTGcCCcGACcCCcCTcCGAA 6114
QY 2000 sAsnLeuAlaProHISerHISerAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
DB 6115 GAACcCTcCGACcCTcACcCAcCGcAGcCGcCGcAGcCGcCGcCGcACcCTGcCGcCTcCGG 6174
QY 2020 pProHISerArgGlySerThrGlnSerIleProPheSerIleGlnGluLeuGluLeuArgSe 2040
DB 6175 CcCGcGACcCGGGAAGAcCTCAAGcCTTcCTTcCATcCGcAGAAcCTGAAcCTcCGTTC 6234
QY 2040 rLeuGlyTYrHISglYSerSerTYrSerProGluGlyValGluProValSerProValSe 2060
DB 6235 TcTGAGGTcCACcGAGcAGcACTcACcCCcCGAAAGGgGTGAGcCCcGTACcGCTcGTAG 6294
QY 2060 rSerProSerLeuThrHISAspLySglYLeuProIlyHISLeuGluLeuAAspLySe 2080
DB 6295 CTACcCACTcGTAGcACcCAcGAGcAGcGGcCTCCcCAAGcACcCTGcAGAAcGACTcGACAAAG 6354
QY 2080 rHISLeuGluGlyGluLeuArgProIlySglInProGlyProValTySLeuGlyGlyGluAl 2100
DB 6355 CCAcCTGcAGcGGGgAGcCTcCGcCCcCAAGcAGcAGcCGcCGcGTGAAcCTTGTGGcGGGAGG 6414
QY 2100 sAlaHISLeuProHISLeuArgProLeuPProGluSerGlnProSerSerSerProLeu 2120
DB 6415 CcGcCCcACTcCCcACcACTcCGcCGcGTGcCTGcAGACcAGcCCcCTcGTCCcAGcCCcGTGCT 6474
QY 2120 uGlnThrAlaProGlyValIlySglYHISglnArgValIleThrLeuAlaGlnHISerIleSe 2140
DB 6475 CcAGACcCGcCCcAGcGGGTcCAAGGTcACcAGcGGGgGTGCACcCTGcCCcAGcACATcAG 6534
QY 2140 rGluValIleThrGlnAspTYrThrArgHISerProGlnGlnLeuSerAlaProLeuPr 2160
DB 6535 TGAAGTCATcACcAGcAGcCTACcACcCGcGcACcCAcCGcAGcAGcCTcAGcCGcACcCTcG 6594
QY 2160 sAlaProLeuTYrSerPheProGlyAlaSerCySProValleuAAspLeuArgArpProPr 2180
DB 6595 CcGcCCcCTcTACTcCTcCTcGTGGGcCAAGcTGCcCCcGTcCTGAGcCTcCGcCGcCCcAC 6654
QY 2180 oSerAAspLeuTYrLeuProProProAspHISglYAlaProAlaArgGlySerProHISer 2200
DB 6655 CAGTGAcCTcTACcCTcCGcCCcCGcGACcCATcGTGTcCGcCGcCGcGTcCCcCCcCAG 6714
QY 2200 rGluGlyGlyLySArgSerProGluProAsnLySThrSerValleuGlyGlyGlyLys 2220
DB 6715 CcGAAGcGGGcCAAGAGGTcCTcCAAGAGcCAAAAGcGTcGTGGGTGTGGGTGAGGA 6774
QY 2220 pGlyIleGluProValSerProProGluGluYMetThrGluProGlyHISerArgSerAl 2240
DB 6775 CcGTATcGAACcCTcGTcCTcCCcACcCGAGGcCATcGAGcAGcCGcAGcACTcCCcGGAGTGC 6834
QY 2240 aValTYrProLeuLeuTYrArgAspGlyGluGlnThrGluProSerArgMetGlySerIly 2260
DB 6835 TGTGTACcCGcGTcGTcTACcCGGAGATGGGAGACAGcAGcCGcCAcGAGcAGATGGGTCCAA 6894
QY 2260 sSerProGlyAAsnThrSerGlnProProAlaPhePheSerIlySLeuThrGluSerAAsnSe 2280

[illegible]

XX (GENTH) GENENTECH INC.
PA (WUTD/) WU T D.
PP (ZHOU/) ZHOU Y.
XX
PI Wu TD, Zhou Y;
XX WPI, 2004-534300/51.
XX
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 1338, 5504pp; English.

The present invention describes an isolated tumour-associated antigenic target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of (a); (c) the complement of (a) or (b); (d) a sequence that has 80% sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-(c). Also described: (1) an expression vector comprising the above nucleic acid; (2) a host cell comprising the above expression vector; (3) a process for producing a polypeptide; (4) an isolated polypeptide comprising: (a) an amino acid sequence encoded by any of the above nucleotide sequences; (b) an amino acid sequence encoded by the full-length coding region of the above nucleotide sequences; or (c) a sequence having at least 80% identical to (a) or (b); (5) a chimeric polypeptide comprising the above polypeptide fused to a heterologous polypeptide; (6) an isolated antibody that binds to the above polypeptide; (7) a process for producing the antibody; (8) an isolated oligopeptide that binds to the above polypeptide; (9) a tumour-associated antigenic target (TAT) binding organic molecule that binds to the above polypeptide; (10) a composition of matter comprising the above (chimeric) polypeptide, antibody, oligopeptide or TAT binding organic molecule, in combination with a carrier; (11) an article of manufacture comprising a container and the composition of matter contained within the container; (12) methods of inhibiting the growth of a cell that expresses the above protein, where the growth of the cell is at least in part dependent upon a growth potentiating effect of the above protein; (13) a method of therapeutically treating a mammal having a cancerous tumour comprising cells that express the above protein; (14) a method of determining the presence of a protein in a sample suspected of containing the protein described above; (15) methods of diagnosing the presence of a tumour in a mammal; (16) a method for treating or preventing a cell proliferative disorder associated with increased expression or activity of the above protein; and (17) a method of binding an antibody, oligopeptide or organic molecule to a cell that expresses the protein described above. The TAT sequences have cytostatic activities, and can be used in gene therapy. The composition and methods are useful for diagnosing, preventing or treating cancer. The composition is also used for preparing a medicament for the therapeutic treatment or diagnostic detection of a cell proliferative disorder or cancer. The present sequence represents a human TAT cDNA sequence from the present invention.

Sequence 8533 BP, 1836 A; 3034 C; 2528 G; 1135 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Length:
Score: 13119.50	Matches: 8533
Percent Similarity: 99.56%	Conservative: 2506
Best Local Similarity: 99.52%	Mismatches: 1
Query Match: 99.28%	Indels: 5
DB: 13	Gaps: 1

US-09-522-753-5 (1-2517) x ADQ84524 (1-8533)

OY MetSerGIserThrGlnLeuValAlaGlnThrTPARGalAthrGluProArgTYrPio 20
Db 2 ATGTGGGATCCACACACAGCTGTGCACAGACTGGAGGCCACTATAGCCCGCTACC 61

OY 21 ProHisSerLeuSerTYrProValGlnIlealaArgThrHisTrAspValGlyLeu 40
Db 62 CCCACAGACCTTTTCCTACCCAGTGCAAGATCGCCCGCAGCACACGGAAGCTCGGGCTCTC 121

QY 41 GltYrGlnHnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 122 GAGTACACACACACACCTCCGCGACATATGCTCCCACTGTCCGCGCTCCATCATCCAG 181
QY 61 ProGlnArgArgProSerIleuSerGlnPheGlnProGlyAsnGlnArgSerGln 80
DB 182 CCCGAGCGGAGGAGGCTCTCCCTGCTGTCTGAGTTCCAGCCCGGAGATAGAACGCTCCAG 241
QY 81 GltLeuHisLeuArgProGlySerHisSerTyrLeuProGlyIleuGlyIleuSerGlnMet 100
DB 242 GAGCTCCACCTGGCGGACAGTCCCACTACCTCCGAGGTGGGAGTAGAGTAG 301
QY 101 GltPheIleGlySerIleArgProArgLeuGlyIleuLeuProAspProLeuLeuArgPro 120
DB 302 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTCTGCTGACCCCTGCTGCGACCG 361
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlyAspLeuThrIleArgAspArgSer 140
DB 362 TCACCCCTGCTGGCCACGGGACGCTGGGGATCTGAAGACTTCACCAAGACCGTATGC 421
QY 141 LeuThrGlyIleLeuGlnProValSerProProSerProProHisThrAspProGlyLeu 160
DB 422 CTGACCGGAGAGTGGAAACCGGTGTCTCCCGACGCCCGCCGACACTGACCTGAGCTG 481
QY 161 GltLeuValProProArgLeuSerIleArgGlyIleuIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGTGCTCCGCGACGCTGTCCAGAGAGAGCTGATCCAGAACATGAGCCGCTGAGAC 541
QY 181 ArgGlyIleThrMetValGlyGlnGlnIleSerIleLeuIleIleValIleArgGlnGln 200
DB 542 CAGAGATACCACTGGTATAGAGCAGCAGATCTTAACTGAAAGAAAGCAGCAACGCTG 601
QY 201 GltGlnGlyAlaAlaIleArgProProGlyIleuProGlyIleuProValSerProProIleGln 220
DB 602 GAGAGGAGAGGCTGCCAAGCCGCCGAGCTGAGAAACCCGTGTCAACGCCGCCATCGAG 661
QY 221 SerIleHisIleArgSerIleuValGlnIleIleTyrAspGlnAsnArgIleValIleGln 240
DB 662 TCGAAGCACCGCAGCTGTGTGAGATCATCTACGAGAACCGGAAAGAGGTGAGAGCT 721
QY 241 AlaHisArgIleLeuGlnGlyIleuGlyIleuProGlnIleuValGlyLeuProIleuThrArgGlnIlePro 260
DB 722 GACATCGGAGATCTGGAAGGCTGGGGGCCCGAGGTGAGGTGCGCTGTACCAACGAGCCC 781
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleIleValIleAsnGlnAlaMetArgIleValIleu 280
DB 782 TCCGACACCCGCGAGATCATGAGAACATCAAAATAAACAGGCGATGCGGAAAGAGCTA 841
QY 281 IleLeuTyrPheIleArgIleArgAsnHisAlaIleArgIleGlnTyrIleGlnIlePheCysGln 300
DB 842 ACTTGTACTTCAAGAGAGGAATCACGCTCGGAAACAATGGAGAGAGAAAGTTCTCCAG 901
QY 301 ArgTyrAspGlnIleuMetGlnAlaLeuGlyIleuValGlyIleuValIleGlnAsnAsnPro 320
DB 902 CGGTATGACAGCTCATGAGGCTGTGGAGAAAGAGTGGAGCATTCAGAAACAACCCC 961
QY 321 ArgArgArgAlaIleArgIleuSerIleValArgGlyTyrTyrGlyIleuSerGlnPheProGlyIle 340
DB 962 CCGCGCGCGGCGCAAGAGAGAGAGAGTGGCGGAGTACTACAGAGAGAGAGTTCCTCGAGATC 1021
QY 341 ArgIleGlnIleArgIleuLeuGlnIleuArgMetGlnSerArgValGlyIleuArgIleSerGly 360
DB 1022 CCCAAGCACCGCAGCTGAG 1081
QY 361 LeuSerMetSerIleAlaIleArgSerGlnHisGlyIleuValSerGlnIleIleAspGlyLeuSer 380
DB 1082 CTGTCCATGTGGCGCCCGCGACGAGCACAGAGTGTTCAGAGATCATTCAGTGGCTCTCA 1141
QY 381 GltGlnGlnIleAsnLeuGlyIleuArgIleMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1142 GAGCAG 1201
QY 401 AspAlaAspGlnGlnIleArgIleIleValIleValIleAsnMetAsnGlyLeuMetAlaAspProMet 420

DB 1202 GACGCTGACACACACGACATCAAGTTCAATCAACATGAAGGGCTTATGGCCACACCCATG 1261
QY 421 LysValIleTyrLysAspArgGlnValIleMetAsnMetTPSerGlnGlnIleuIleValIleThrPhe 440
DB 1262 AAGGTGTACAAAGACCGCGAGTCTATGAACATGTGAGTGAACAGAGAGAGAGAGAGAGAG 1321
QY 441 ArgGlyIlePheMetGlnHisProIleAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1322 CCGGAGAAAGTTGATGACAGATCCCAAGAACTTTGGCTGTATGATCATTTCTCTGGAGAGG 1381
QY 461 LysThrValAlaGlyCysValLeuTyrTyrTyrIleuThrIleValAsnGlnIleuAsnTyrLys 480
DB 1382 AAGACAGTGGCTGAGTGTGCTCTTATTTATCTGACTGAAGAAGATAGAACTATATAG 1441
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyIleValSerGlnGlnIleuGlnIleuGln 500
DB 1442 AGCTGTGTAGACGAGAGCTATGTGGCGCCCGGCAAGGACACAGACAGACAGACAGAG 1501
QY 501 Gln 520
DB 1502 CAG 1561
QY 521 AspGlyIleValIleGlyIleuValIleGlyIleuValIleGlyIleuValIleGlyIleuVal 540
DB 1562 GATGAGAAAG 1621
QY 541 AspLysGlnAspLeuLeuIleuValIleValIleThrAspAspThrSerGlyIleuAspAsnAspGlu 560
DB 1622 GACAAAGAAAG 1681
QY 561 LysGlnAlaValAlaSerIleValIleValIleValIleValIleValIleValIleValIle 580
DB 1682 AAGAGAGCTGTGGCTCCAAAGCGCGCAAAAGCTGCAACAGCAGAGAGAGAGAGAGAGAG 1741
QY 581 ArgIleThrArgSerMetIleAsnGlnIleAsnSerGlnAlaIleThrProGlnGln 600
DB 1742 CGCATCACCGCTCAATGCTATATGAGCACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
QY 601 SerAlaGlnLeuAlaSerMetGlnLeuAsnGlnIleuSerSerArgTyrThrGlnGlnIleuMet 620
DB 1802 AGCGCGAGCTGGCTCCATGAGCTGAAATGAGATGATTCGCTGAGCAGAAAGAGAAATG 1861
QY 621 GltThrAlaIleLysGlyIleuLeuGlnHisGlyIleArgMetTPSerAlaIleAlaArgMet 640
DB 1862 GAAACAGCCAAAGAGTCTCTGGAACAAGCGCCGCAATGATGTGGCTCATTCGCCCGAGAG 1921
QY 641 ValGlySerIleThrValSerGlnCysIleAsnPheTyrPheAsnTyrIleLysArgGln 660
DB 1922 GTGGGCTCCAAAGCTGTGTGAGAGTGAAGACTTCTTCAACTACAAAGAGAGAGAGAG 1981
QY 661 AsnLeuAspGlnIleuGlnGlnIleHisIleLysLeuIleMetGlnLysGlnIleArgAsnIleArg 680
DB 1982 AACCTCGATGATCTTTCAGACAGACAGAGCTGAAAGTGAAGAGAGAGAGAGAGAGAGAG 2041
QY 2042 AGGAG 2101
QY 701 GltAspGlnGlnIleuMetGlnAlaSerGlyValIleSerGlyIleuGlnIleuValIleGlnGlu 720
DB 2102 GAGGATGAG 2161
QY 721 AlaGlnAlaLeuHisAlaSerGlyIleuGlnIleuValProArgGlyIleuIleuSerGlyProAla 740
DB 2162 GCTGAAGCTTCAATGCTCTGTGGAAATGAGTGTCCCAAGGGGAGATGCAATGGCCCAAGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisThrGlnAlaAlaLys 760
DB 2222 ACTGTCAACAACAGCTCAAGACACGAGAGAGATCCCTCTCTCAACACTGAGGCGCGCAAG 2281
QY 761 AspThrGlyIleuAsnGlyProIleProAlaThrIleuGlyValAspGlyProProPro 780

[illegible]

QY	1140	gAlaIaYValaAProValGIYProValThrMetGlyLeuProLeuProMetLarProIyValY	1160
Db	3421	TGCAAGGAGCCCGGGGGGCGCTGTCAACATGAGGGGTGCCTGTGCCATGAGACCCCAAAA	3480
QY	1160	sLeuAlaProPheserGIYValIyGSGInGIleuLeuSerProArgIyGIaIaGIYPr	1180
Db	3461	GCTGGACACCTTTCACCGGAGTGAAGACGAGACACTGTCCCAAGGGCCAGGCTGGCC	3540
QY	1180	oProGIuSerLeuGIYValProThraIaGIuIaIaSerValLeuArgIYThraIaIe	1200
Db	3541	ACCGAGAGGCTGGGGGGTGGCCCAAGGCCCAAGAGGGCGTCCGTGTGAGAGGAGACACT	3600
QY	1200	uGIYSerAlaProGIyGIYserIleThrIyGIYIleProSerThraYValProSerAs	1220
Db	3601	GAGCTCACTTCCGGCGGAGACATCCAAAGGCAATTCACAGACACGGAGGCTCCCGA	3660
QY	1220	pSerAlaIleThrTYaArgIYserIleThraIGIYThProAlaAraValLeuTYLY	1240
Db	3661	CAGCGGCATCAATACCGGGCTCATCATCAACCAAGGACGGCAGCTGACTGTACAA	3720
QY	1240	sgIYThrIleThraArgIleIleGIYGIuAraSerProSerArgLeuAraArgIYArgGI	1260
Db	3721	GGGACCATCAACCAAGATCATCGGCGAGAGACAGCCGATGCTTGACCGCGGCGGGA	3780
QY	1260	uAraSerLeuProArgGIYsValIleTYGIuGIYLYaYsGIYhIaValLeuSerTY	1280
Db	3781	GGAACGCTGGCCCAAGGGCCACGTCATCTAGAAAGCAAGAAAGGCCACGCTTGTCTTA	3840
QY	1280	rGIuGIYGIYMeSerValThrGIYSerIyGIuAraArgIYArgSerSerSerGIYPr	1300
Db	3841	TGAAGGTGTGACATGTCTGTAGACCCAGTCTCCAAAGAGAGACGGCAAGACCTCAGACC	3900
QY	1300	oProhIGIuThraIaIaProIyAraGTrThTYAraMetMetGIuGIYArgValIGIYAr	1320
Db	3901	CCCCCATGAGACGGCGCGCCCAAGCGCACTTAAGACATGATGAGAGGGCGCGTGGCAG	3960
QY	1320	gAlaIleSerSerAlaSerIleGIuGIYLeuMetGIYArgAlaIleProProGIuArgAl	1340
Db	3961	AGGCATCTCTCCAGCCAGCATCGAAGGTCTCATGGCGCGTGCATCCGCGGAGCGACGA	4020
QY	1340	sSerProhIhIaIeIyGIuGIuThraIaIeArgIYserIleThrGIuGIYIlePr	1360
Db	4021	CAGCCCCACCATCTCAAGAGAGACGACCAATCCGGGGCTCATCAACAAAGGAAATCC	4080
QY	1360	oArgSerTYaGIuIaGIuIaGIuAraArgTYLeuAraArgIuIaIaIyLeuLeuIyAra	1380
Db	4081	TGGTCTCTACGTGAAGGACAGAGAGACTACGTGCGTGGGAAGGCCAAAGCTCTAAAGCG	4140
QY	1380	gGIuGIYThrProProProProProProSerAraArgAraLeuThrGIuIaTYIYThrGI	1400
Db	4141	GGAAGGACAGCTTCGCGCCACCGCCCTCAAGGAACTGAACCAAGGCTCAACAAAGACGA	4200
QY	1400	nAlaLeuGIYProIeIyLeuIySProAlaIhIGIuGIYLeuValIaIaThrValIyArgI	1420
Db	4201	GAGCTTGAGCCCTGAAAGCTGAAGCGGCGCCATGAAGGGCTGGTGGCACAGGTGAAGGA	4260
QY	1420	uIaGIYArgSerIleIhIGIuIleProArgIuGIuLeuArgIhIaThrProGIuLeuPr	1440
Db	4261	GAGCGGCGCGCTCATTCATGAGATCCGCGCGAGAGAACTGGGACACAGCCCGAGCTGCC	4320
QY	1440	oLeuAlaProArgProIeIyGIuGIYserIleThrGIuGIYThrProIeIyTYIYAs	1460
Db	4321	CTGGGCCCCGGCGGCTCAAGAGGGCTCATCAACCAAGGACCCCGCTCAAGATGAGA	4380
QY	1460	pThrGIYAlaSerThrThrGIYserIySValhIaAraValArgSerLeuIleGIYSerPr	1480
Db	4381	CACCGGCGGCTCAACCATGGCTCAAAAAGACAGATGAGCTCCCTCATCGGACAGCC	4440
QY	1480	oGIYArgThrPhaProProValhIaProLeuAraValMeAlaAraAlaArgAlaIeGI	1500
Db	4441	CGGCGGAGACGTTCACACCGCTGACCGCGCTGATGTGATGGCGAGCGCCGGGACCTGGA	4500

QY 1500 uArGlaCyTgYrGluSerLeuSerLeuSerArGProGlyThrAlaSerSerSerGlyGly 1520
| | | | |
Db 4501 ACtGCGCTGCTCAAGAGAGAGCTGAAGAGCCGGGACCGGAGCAGCTCGGGGGG 4560
QY 1520 ySerTleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPr 1540
| | | | |
Db 4561 CTCACATGGCGCGGGGCGCCGGTCAATTGGCTGAGCTGGGTAAACCCCGGAGAGGCC 4620
QY 1540 oLeuThrTYrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVa 1560
| | | | |
Db 4621 CcTGACCTATGAGAGACACAGGGGACCTTTGGCGGCACTCCACAGAGTTCCGCCCT 4680
QY 1560 lThrMetArGluProThrProArgLeuGlnGluGlySerLeuSerSerSerIysAlaSe 1580
| | | | |
Db 4681 GACCAAGCGGGAGCCACCGCGGCTGAGAGAGGAGCAGCTTTGCTCCAGCAAGGCAATC 4740
QY 1580 rGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrVa 1600
| | | | |
Db 4741 CCAGGACCGAAAGCTGACGTGACCGCTGTGAGATGCGCAAGTCCCGCAGACAGCCGT 4800
QY 1600 lProGluHisSerProHisSerProIleSerProTYrGluHisLeuLeuArgGlyValSerGly 1620
| | | | |
Db 4801 GCCCAGACACACACACACACCCCATCTCGCCCTATGACACCTGCTTCGGGGCGTGAAGTG 4860
QY 1620 yValAspLeuTYrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
| | | | |
Db 4861 CGTGGACCTGTATCGAGCCACATCCCGCTGGCGCTTCGACCCACCTCATACCCCGCG 4920
QY 1640 yIleProLeuAspAlaAlaAlaTYrTYrLeuProArgHisLeuAlaProAspProth 1660
| | | | |
Db 4921 CATCCCTCTGAGAGCGAGCGCGCTGCTACACCTGCGCCACACCTGACCTGACCCCAACCCAC 4980
QY 1660 rTYrProHisLeuTYrProProTYrLeuIleArgGlyTYrProAspThrAlaAlaLeuGly 1680
| | | | |
Db 4981 CTACCCGACCTGTACCCACACCTCATCTCCGGCTACCCCGACACGGCGCGCTGGA 5040
QY 1680 uAsnArGlnThrIleIleAsnAspTYrIleThrSerGlnGlnMetHisIleAsnThrAl 1700
| | | | |
Db 5041 GAACCGGACAGACCATCATCATACATCACTCGAGCAGATGACACACCAACCGGCG 5100
QY 1700 aThrAlaMetAlaGlnArGlaAspMetLeuArgGlyLeuSerProArgGlnSerSerIe 1720
| | | | |
Db 5101 CACCGGCATGCGCCACGAGCTGATGTGAGAGGGCGCTCTCCCGCGAGTCCTCGCT 5160
QY 1720 uAlaLeuAsnTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisSe 1740
| | | | |
Db 5161 GGCACTCAACTACGCTGCGGGTCCCGAGGATCATCGACTTCCCAAGTCACACCT 5220
QY 1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
| | | | |
Db 5221 GCCTGTGCTCGTCCCGCCGACACAGGACCCACGACCGCATGGAACCGCTTGCTGA 5280
QY 1760 rLeuProThrAlaProGlnProPheSerSerArGHisSerSerSerProLeuSerProGly 1780
| | | | |
Db 5281 CcTCCCGACCGCGCCCGACGCTTCAGCAGCGCCACAGCAGCTCCCACTTCCCGCAGG 5340
QY 1780 yGlyProThrHisLeuThrIysProThrThrThrSerSerSerGlyAlaArgGlyAspAr 1800
| | | | |
Db 5341 AGGTCCAAACACTTGAACAAACCAACACACAGCTCTCTGTCGAGCGGAGGAGACCG 5400
QY 1800 gAspArGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrTh 1820
| | | | |
Db 5401 GGATCGAGAGCGGGACCGGGATCGGGAGCGGAAAGTTCATCTCCACGTCACACAGAC 5460
QY 1820 rValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySe 1840
| | | | |
Db 5461 GGTGGAGCAGCAGCCCATCTGAGACCTGTGTACAGAGCAG-----AGCAGCGGAG 5511
QY 1840 rSerGlyGlyGlyGlyGlySerSerSerArGProAlaSerHisSerHisAlaHisGlnHis 1860
| | | | |
Db 5512 CACCGGGGGGGGGGGGAGCAGACCGCCCGGCTCCCACTCCCATCCACACAGCA 5571
QY 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAs 1880

Db 5572 CTGCGCCCAATCTCCCTCGGACCCAGAGATGCTCTCCAGCAGAAACCAAGTCTTCACAA 5631
QY 1880 nThrGlyMetIysGlyIleIleThrAlaValGluProSerIysProThrValLeuArgSe 1900
| | | | |
Db 5632 CACAGGATGAAGGGTATCATACCGCTGTGAGGCCAGCAGCCACCGTCTTGAGGTTC 5691
QY 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCyPr 1920
| | | | |
Db 5692 CACTTCACCTCTCAACCGGTTCGCGCGGCTCCCATTCCTCCACCTGCGCACCCACTGCC 5751
QY 1920 oLeuGlyGlyThrIleuAspGlyValTYrProThrIleuMetGluProValLeuLeuProLy 1940
| | | | |
Db 5752 ACTGGCGGACCCCTGATGGGGGTCTACCTTACCTCATGAGGCCGTTCGTGCTGCCAA 5811
QY 1940 sGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLe 1960
| | | | |
Db 5812 GGAGGCCCGCCGGGTGCGCCGCGCAGAGGGCGCCGAGCAGACCGGCGCATGCTTCTCT 5871
QY 1960 uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerIysSerGly 1980
| | | | |
Db 5872 CCGCAAGCCCCCAGCGCGCTCGGGCTGAGCGCGCTCTCCCTCCCGCAGAAAGGCTCGGA 5931
QY 1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaTy 2000
| | | | |
Db 5932 GCCCGGCGCCCTAGTGTCTCTGTCTGTGCGCACGACCATGCGGACCCCTCGAA 5991
QY 2000 sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
| | | | |
Db 5992 GAACCTCGACCTTCACACACAGCCAGCCCGGACCGCGCGCCACGCTCGGCGCTCGGA 6051
QY 2020 pProHisArgGlyLysThrGlnSerIysProPheSerIleGlnGluLeuGluLeuArgSe 2040
| | | | |
Db 6052 CCGGACACCGGAAAGACTCAAGTAACCTTTTCATCCAGCAAGAACTCGGTTTC 6111
QY 2040 rLeuGlyTYrHisGlySerSerTYrSerProGluGlyValGluProValSerProValSe 2060
| | | | |
Db 6112 TCTGGATTACAGGAGGAGCTACAGCCCGCAAGGGGGTGAAGCCGTGAGCCCTTGAG 6171
QY 2060 rSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSe 2080
| | | | |
Db 6172 CTACACCACTGTGACCCACAGACAGAGGGGCTCCCAAGCACTGGAAAGACTGACAAAG 6231
QY 2080 rHisLeuGluGlyGluLeuArgProIysGlnProGlyProValLysLeuGlyGlyGluAl 2100
| | | | |
Db 6232 CCACCTGAGGGGAGACTCGGCGCCCAAGCAGCGCCCGTGAAGCTTGCGCGGAGAGG 6291
QY 2100 aAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuSe 2120
| | | | |
Db 6292 CGCCCACTTCACACACTCGCGCGCTGCTTGAAGACCGCTCTGTCCAGCCCGCTGCT 6351
QY 2120 uGlnThrAlaProGlyValIysGlyHisGlnArgValValThrLeuAlaGlnHisIleSe 2140
| | | | |
Db 6352 CCAAGCCGCCCAAGGGGTCAAAAGTCAACAAGGGGTGTCACTGTGCCAGACATCAAG 6411
QY 2140 rGluValIleThrGlnAspTYrThrArgHisHisProGlnGlnLeuSerAlaProLeuPr 2160
| | | | |
Db 6412 TGAGGTCAATCAACAGAGTCAACCCGGCACCACCAAGCAGCTTCAAGCGCACCTCGCC 6471
QY 2160 oAlaProLeuTYrSerPheProGlyAlaSerCySPProValLeuAspLeuArgProPr 2180
| | | | |
Db 6472 CGCCCCCTTACTCTCTTCCTGGGGCAGCTGCCCGGTCTGAGCTTCGCGGCCACCC 6531
QY 2180 oSerAspLeuTYrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSe 2200
| | | | |
Db 6532 CAGTGACTCTACTCTCCCGCCCGGACCAATGTGCCCCGGCGCGTCCCCCACAAG 6591
QY 2200 rGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyLys 2220
| | | | |
Db 6592 CGAAGGGGAGAGGTCTCCAGAGCCAAAGAGTGTGTGTGGTGTGTGTAGGA 6651
QY 2220 pGlyIleGluProValSerProProGluGluMetThrGluProGlyHisSerArgSerAl 2240
| | | | |

D 6652 CGGATTTGAACCTGTGTCCTCCACCGGAGGCGATGACGAGCCAGGCGCATCTCCGAGTGC 6711
Q 2240 aValTyrProleuLeuTyrAArgAspGlyGluGlnThrGluProSerAArgMetGlySerly 2260
D 6712 TGTGTACCCGCTGCTGTACCGGAGTGGGAAACAGACGAGCCCGACAGAGATGGGCTCCAA 6771
Q 2260 sSerProGlyAAsnThrSerGlnProProAlaPhePheSerlyLeuThrGluSerAAsn 2280
D 6772 GTCTCCAGGCAACACGAGCCAGCCGCGCTTCTTTCAGCAAGCTGACCGAGCAACTC 6831
Q 2280 rAlaMetVallySerlySerlyGlnGluIleAsnlyLeuLeuAsnThrIleAsnAArgAs 2300
D 6832 CGCCATGGTCAGAGTCCAGAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCA 6891
Q 2300 nGluProGlyUTrAAsnIleSerGlnProGlyUTrGlnIlePheAsnMetProAlaIleTh 2320
D 6892 TGAGCTGATACATATTCAGCCAGCTGGGACGAGATCTTCAATATGCCCCGATAC 6951
Q 2320 rGlyThrGlyLeuMetThrTyrAArgSerGlnAlaValGlnGluIleAsnIleSerThrAsn 2340
D 6952 CGGAAACAGGCTTATGACTATAGAGCAGCGGTGAGAGACATGCGACAGCAACAT 7011
Q 2340 rGlyLeuGluAlaIleIleArglyAlaLeuMetGlyUTrGlyTyrAspGlnTrpGluGlu 2360
D 7012 GGGGCTGAGGCGCATTAATTGAAAGGCACTCATGGGTAAATATGACAGTGGGAAAGTTC 7071
Q 2360 rProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerIleSerLeuProAlaI 2380
D 7072 CCGCCGCTGACCCGCAATGCTTTAAACCTCTAAGTCCAGTCCAGCTGCGCCGCTGC 7131
Q 2380 aMetProIleThrAlaIleAspGlyAArgSerAspIleThrLeuThrSerProGlyGly 2400
D 7132 TATGCCATTAACCGCTGCTGACGAGCGAGAGACACACATCACTCGCCAGGTGGCGG 7191
Q 2400 yGlyUTrAlaIleValSerGlyAArgProSerSerAArglyAlaUTrSerProAlaProG 2420
D 7192 CGGGAAGGCGCAAGTCTCTGCGAGACCGACGAGCGGAAAGCGCAAGTCCCGCGCGG 7251
Q 2420 yLeuAlaSerGlyAAspAArgProProSerValIleSerValIleSerGluGlyAspGlyAs 2440
D 7252 CTTGGCATCTGGGAGCGGCGACCTCTGCTCTCTGATGACTCGAGGAGGAGCTCCAA 7311
Q 2440 nArgArgThrProLeuThrAAsnAArgValIleGluAspAArgProSerSerAlaGlySerTh 2460
D 7312 CCGCGGAGCGCGCTGACCAACCGCGTGGGAGAGACAGCGCTCCGACAGGTTCCAC 7371
Q 2460 rProPheProTyrAAsnProLeuIleMetAArgLeuGlnAlaGlyValIleAlaSerProPr 2480
D 7372 GCCATTCCTCTACAAACCCCTGATCATGCGGCTGACGCGGCTGTCAATGCTCCAC 7431
Q 2480 oProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProIleIleAlaTrpAspG 2500
D 7432 CCGACCGGCGCTCCCGCGGAGCGGAGCGGCGCTGCTGCGCCCGACACCGCTGGAGCA 7491
Q 2500 uGluProLyseProLeuLeuCyseSerGlnTyrGluThrLeuSerAAspSerGln 2517
D 7492 GAGGCCCAAGCACTGCTGCTGCTGCGAGTACGACACACTCTCCAGACGCGAG 7543

RESULT 7

ACN39603 standard; CDNA: 8533 BP.
ACN39603;
ACN39603;
18-NOV-2004 (first entry)
Tumour-associated antigenic target (TAT) cDNA DNA325805, SEQ ID NO:3856.
Tumour-associated antigenic target; TAT; human; overexpression; cancer;
tumour; diagnosis; cell proliferative disorder; breast cancer;
colorectal cancer; lung cancer; ovarian cancer; liver cancer;
central nervous system cancer; bladder cancer; pancreatic cancer;
cervical cancer; melanoma; leukemia; hybridisation probe;

KM chromosome identification; chromosome mapping; gene mapping;
KM gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX MO2004030615-A2.
XX 15-Apr-2004.
XX 29-SEP-2003; 2003MO-US028547.
XX 02-OCT-2002; 2002JUS-0414971P.
XX (GERTH) GENENTECH INC.
XX Mu TD, Zhang Z, Zhou Y;
XX WPI; 2004-347921/32.
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX Claim 1; SEQ ID NO 3856; 7273bp; English.

XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acid and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid, an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention

XX Sequence 8533 BP; 1836 A; 3034 C; 2528 G; 1135 T; 0 U; 0 Other;

Alignment Scores:

XX Pred. No.: 0 Length: 8533
XX Score: 13119.50 Matches: 2506
XX Percent Similarity: 99.56% Conservative: 1
XX Best Local Similarity: 99.52% Mismatches: 7
XX Query Match: 99.28% Indels: 5
XX DB: 13 Gaps: 1

US-09-522-753-5 (1-2517) X ACN39603 (1-8533)

Q 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProAArgTyrPro 20
D 2 ATGTGGGATCCACACAGCTGTGGACAGACAGCTGGAGGGCCACTGAGCCCCGTACCGG 61
Q 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrIleThrAspValAlaGlyLeu 40
D 62 CCCACAGGCTTCTTCAACCCAGTGCAGATGCGCGGACGACACGAGCGTGGGCTCTCG 121
Q 41 GluTyrGlnHisIleSerAArgTyrAlaSerHisIleUTrSerProGlySerIleIleGln 60
D 122 GATTAACAGACACACTCCCGGACTATGCTCCCACTGTGCGCGCTCAATCATCAG 181
Q 61 ProGlnArgArgArgProSerLeuLeuSerGlnPheGlnProGlyAAsnGluArgSerGln 80

Db 182 CCCAGCGGCGAGGCCCTCCCTGCTGTGAAGTCCAGCCGGGAATGAACGGTCCAG 241
Qy 81 GJuLeuH1sLeuAArgProGluSerH1sSerTyrluProGluLeuGlyLysSerGluMet 100
Db 242 GAGCTCACCTCGGCGCAAGAGTCCCACTATACCTGCGGAGCTGGGGAAGTCAGAGATG 301
Qy 101 GJuPhel1eGluSerLysArgProAArgLeuGluLeuLeuProAArgProLeuLeuAArgPro 120
Db 302 GAGTATATTGAAGCAAGCGCCCTCGGCTGAGAGCTGCTGCTGACCCCTGCTGCCACCG 361
Qy 121 SerProLeuLeuA1aThrGlyGlnProA1aGlySerGluAAspLeuH1rLysAspAArgSer 140
Db 362 TCAACCTCGCTGGCCACGGGCGAGCCCTGGGGATCTGAAGACCTCAACAAAGACCGTAGC 421
Qy 141 LeuThGlyLysLeuGluProVal1SerProProSerProProH1sThraArgProGluLeu 160
Db 422 CTGACGGGCAAGTGAACCCGATGTCTCCCCCAAGCCCCCGCAACATGACCTTGAGCTG 481
Qy 161 GJuLeuVal1ProProAArgLeuSerLysGluGluLeu11eGluAAspMetAAspArgVal1Asp 180
Db 482 GAGCTGTGCGCCGACAGCTGTCCAAAGAGAGCTGATCCAGAACTGGACCGCGTGAGAC 541
Qy 181 ArgGlu11eThraMetVal1GluGlnGln11eSerLysLeuLysLysGlnGlnGlnLeu 200
Db 542 CGAGAGATCAACATGATGAAGCAGACAGATCTTAAGCTGAAGAAAGCAGCAACAGCTG 601
Qy 201 GJuGluGluVal1aA1aLysProProGluProGluLysProVal1SerProProPro11eGlu 220
Db 602 GAGGAGAGAGGCTGCCAAGCGCCCGGAGCTGAGAAAGCCGCTGACCGCGCCCATCGAG 661
Qy 221 SerLysH1sArgSerLeuVal1Gln11e11eTyraSpGluAAspArgLysLysVal1aGluVal1a 240
Db 662 TCGAAGCACCGCAGCTGTGAGATCATCTAAGAGAAACCGGAAGAGCTGAAGCT 721
Qy 241 A1aH1sArg11eLeuGluGluGlyLeuGlyProGluVal1GluLeuProLeuTyraAsnGlnPro 260
Db 722 GCACATCGAATCTGGAAGGCTGGGGGCCCAAGTGAAGCTGCGCTGTAACAACAGGCC 781
Qy 261 SerAAspThraArgGlnTyra1sGluAAsn11eLys11eAAsnGluA1aMetAArgLysLysLeu 280
Db 782 TCCGAACACCGGAGATGATCATGGAACATCAAAATTAACCAAGCGATCGCGAAGAAAGCTA 841
Qy 281 11eLeuTyraPheLysArgArgAAsnH1sA1aArgLysGlnTyraGluLysPheCysGln 300
Db 842 ATCTTACTTCAAGAGAGAGAAATCAAGCTCGAACAATGAGAGCAAGAAATGCTCCAG 901
Qy 301 ArgTyraAspGluLeuMetGluA1aLeuGluLysLysVal1GluArg11eGluAAspAAspPro 320
Db 902 CGCTATGACCAAGCTCATGAGGCTGGGAGAAAGAGTGAAGCGCATCGAAGAAACACCC 961
Qy 321 ArgAArgA1aLysGluSerLysVal1ArgGluTyraTyraGluLysGlnPheProGlu11e 340
Db 962 CGCGCGCGGCGCAAGAGAGCAAGTGGCGGATCTACAGAAAGCAAGTTCCCTGAGATC 1021
Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgVal1GlyLysArgLysGly 360
Db 1022 CGCAACAGCGGAGCTGAGAGAGCGCATGCAAGACAGGGTGGCCACGCGGCAAGTGG 1081
Qy 361 LeuSerMetSerA1aA1aArgSerGluH1sGluVal1SerGlu11e11eAAspGlyLeuSer 380
Db 1082 CTGTTCATGTCCGCCCGCCGACAGCAAGAGTGTCAAGATGATCTGATGCTCTCTCA 1141
Qy 381 GJuGlnGluAAspLeuGluLysGlnMetAArgGluLeuA1aVal11eProProMetLeuTyra 400
Db 1142 GAGCAGAGAACTTGAGAAAGCAGATGCCAGCTGACCTGATCCCGCCCATGCTGTAC 1201
Qy 401 AAspA1aAspGlnGlnArg11eLysPhe11eAAspMetAAsnGlyLeuMetA1aAspProMet 420
Db 1202 GACGCTGACAGAGGAGCATCAAGTTCATCAATGACAGCGCTTATGGCGCAACCCCATG 1261
Qy 421 LysVal1TyrlLysAspArgGlnVal1MetAAspMetTyraSerGluGlnGluLysGluTyraPhe 440
Db 1262 AAGGTGTACAAAGACCGCAGTCAATGAACATGTGAGTGAAGAGAGAGAGACCTTC 1321

Qy 441 ArgGluLysPheMetGlnH1sProLysAAspPheGlyLeu11eA1aSerPheLeuGluAArg 460
Db 1322 CGGAGAAAGTTCATGACAGATCCCAAGAACTTGGCGCTGATGCATCATTCCTGGAGAG 1381
Qy 461 LysThraA1aGluCysVal1LeuTyraTyraTyraLeuThyLysLysAsnGluAAsnTyrlLys 480
Db 1382 AAGACATGAGCTGAGAGCTCTCTTAATTAATCACTGACTTAAGAAAGATGAATGAATTAAG 1441
Qy 481 SerLeuVal1ArgArgSerTyraArgAArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1442 AGCTGTGATGAGAGAGCTATTCGCGCCCGCGCAAGACCAAGCAGCAGCAACACAGCAG 1501
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
Qy 521 AAspGluLysGluLysGluLysGluA1aGluLysGluGlnGluLysProGluVal1GluAAsp 540
Db 1562 GATGAGAAAGAGAAAGAAAGAGCGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
Qy 541 AAspLysGluAAspLeuLeuLysGluLysTyraAAspThraAAspThraSerGlyLysAAspAAspGlu 560
Db 1622 GACAGAGAAAGCTCTCAAGAGAAAGACAGACGACCTCAAGGAGAGAGCAACAGCAGAG 1681
Qy 561 LysGluA1aVal1aSerLysGlyArgLysThraAAspSerGlnGlyArgArgLysGly 580
Db 1682 AAGAGAGCTGTGGCTCCAAAGCGCGCAAACTGCCAAGCAGCAGGAGAGAGAGCAAGAGC 1741
Qy 581 Arg11eThraArgSerMetA1aAAsnGluA1aAAsnSerGluGluA1a11eThraProGlnGln 600
Db 1742 CGCATCAACCGCTCATAGCTATAGGCGCAACAGGAGAGAGCCATCAACCCCGCAGAG 1801
Qy 601 SerA1aGluLeuA1aSerMetGluLeuAAsnGluSerSerArgTyraThraGluGluGluMet 620
Db 1802 AGCGCGAGCTGGCTCTCAAGAGCTGAATGAAGTTCGCTGACACAAAGAAAGAAAG 1861
Qy 621 GJuThraA1aLysLysGlyLeuLeuGlnH1sGlyArgAAspTyraA1a11eAAspMet 640
Db 1862 GAAACAGCAAGAAAGTCTCTCGGAACAGCGCGCAATCGGTGGCGCATCGCGCGGATG 1921
Qy 641 Val1GlySerLysThraLysSerGlnCysLysAAspPheTyraPheAsnTyrlLysLysAArgGln 660
Db 1922 GTGGGCTCCAAAGCTGTGCGAGTGTAAAGACTTCTTCAACATCAAGAAAGAGAGCAG 1981
Qy 661 AAspLeuAAspGlu11eLeuGlnGlnH1sLysLeuLysMetGluLysGluAArgAAspAAspArg 680
Db 1982 AACCTGATGAGATCTTGAGAGCAGCAAGAGCTGAAGATGAAGAGAGAGAGAGAGAG 2041
Qy 681 ArgLysLysLysLysA1aProA1aA1aAAspArgGluA1aA1aPheProProVal1Val 700
Db 2042 AGGAAGAAAGAAAGCGCGCGCGCGCGCGCAGAGAGAGGCTGCATTCGCGCGCGGTG 2101
Qy 701 GJuAAspGluGluMetGluA1aSerGlyVal1SerGlyAAspGluGluMetVal1GluGlu 720
Db 2102 GAGGATGAGAGATGAGAGCGTGGCGCTGAGACCGGAAAGAGAGAGATGTGTGAGAGAG 2161
Qy 721 A1aGluA1aLeuH1sA1aSerGlyAAsnGluVal1ProAArg11eGluCysSerGlyProA1a 740
Db 2162 GCTGAAGCTTCAATGCTCTGGAAGTGAAGTGGCCAGAGGGGAAGAGAGAGAGAGAG 2221
Qy 741 ThraAAspAAspSerSerAAspThraGluSer11eProSerProH1sThraGluA1aA1aLys 760
Db 2222 ACTGTCAACAACAGCTCAACACAGAGACATCCCTCTCTCAACATGAGCGCGCAAG 2281
Qy 761 AAspThraGlyLysAAsnGlyProLysProProA1aThraLeuGlyA1aAAspGlyProProPro 780
Db 2282 GACACAGGCGCAAAATGGCCCAAGCCCGCAGCAGCACTGTGGCGCGCAGAGAGAGAGAG 2341
Qy 781 GJuProProThraProProArgArgThraSerArgA1aPro11eGluProThraProA1aSe 800
Db 2342 GGGCCACCAACCCCAACAGAGAGAGAGATCCGGCC-CCCACTGAGACCAACCCCGGCTTC 2400

QY	800	RGUUAATHRGUVALAETHRTHPRORPProAaLAPROPSerPProSerAaLAPROPr	820
Db	2401	TGAACCCACCGGAGGCCCCCTACGCCCCCAACAGACCCCCCATGCTCTGCACCTCTCC	2460
QY	820	oValValAProLySgUGUUGUyGUGUUGUthThAaAaAaProProValAGUUGU	840
Db	2461	TGTGATCCCCCAAGAGAGAGAGAGAGACCCGACAGCGCCCCCAAGTGAAGAGAGG	2520
QY	840	YGUUGUUGUySProProAaAaAGUUGUleuAaAaAaPThRGUyLYySaaAGUUGU	860
Db	2521	GGAG	2580
QY	860	uProValLySserGUGUyThThGUGUUAAGUUGUyProAaLySGLyYaaAaA	880
Db	2581	GCCCTCAAGAGCGAGTGACCGAGAGAGCCGAGAGAGGCGCGCAAGAGGCAAGACGC	2640
QY	880	AGUUAaAaAGUUAATHRThAAGUUGUyAaAaAaPThRGUyLYySaaAGUUGUySe	900
Db	2641	GGAGAGCGGTGAGGCGAGCGCGAGAGGCGCTCAAGAGCAAGAGAGAGAGAGGAGAG	2700
QY	900	RGUyAaAaAaThThThThAaLySserSerGUYAaProGUAaPProAaPProSerAaA	920
Db	2701	CGGAGAGGCGACCAAGCAAGCAAGAGCTCGGGCGCCCCCAAGACAGACGACTCCAGTCA	2760
QY	920	TCySserAaAaPGLyUaAaPGLyUaAaAGUUGUyGUYyAaPGLyAaAaAaAaAaAaA	940
Db	2761	CTGCAGTGCAGACGAGTGATGAGGCGAGGCGGCGCAAGAACCGGCTGTCTCTCC	2820
QY	940	oAaRGProSerLeuLeuTh	960
Db	2821	AAGGCGGAGGCTCTACCGCCGACGTGGGACCCCGGCGCAATGCTTCAACCCCAAGACC	2880
QY	960	oLeuAaPGLyUaAaPGLyUaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaA	980
Db	2881	ACTGACCTGGAAGAGCGTGAAGAGAGAGCGGCTGCCATCCCCCATTCAGGTACCA	2940
QY	980	aAaAaAaAGUyProProAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaA	1000
Db	2941	AGTCATAGAGCCCCCGGAGAGAGCGAGCTCCCAAGACGAGCTCCCGGAGCCCCCAC	3000
QY	1000	oProProGUAaA	1020
Db	3001	GCCACCGCAAAACCTGAGCGGAGAGAGAGCGCCCTCAGCAGCGCTGACAGAGCCCCG	3060
QY	1020	GGUYySserAaA	1040
Db	3061	GAGGCAAG	3120
QY	1040	nLYaA	1060
Db	3121	GAAAGCTGCTGGGAG	3180
QY	1060	GGUyUaA	1080
Db	3181	TGAGGTGATCAAGAGGCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3240
QY	1080	oGUYyAaA	1100
Db	3241	TGATACCAAG	3300
QY	1100	oTh	1120
Db	3301	CACCATCTCCAAACCGGCTCCCTCATCTCTCCGCAAGAGAGAGAGAGAGAGAGAG	3360
QY	1120	GGUyUaA	1140
Db	3361	GCAATATAGTGCATCTCCCAAGAGATGCGTCAAGCTCCAGCTCCGCTACTCAGAGCA	3420
QY	1140	aAa	1160
Db	3421	TGCCAAG	3480
QY	1160	aLeuAaA	1180
Db	3481	GCTGACACCTTACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3540
QY	1180	oProGUAaA	1200
Db	3541	ACCGAG	3600
QY	1200	UGUYySserAaA	1220
Db	3601	GGCTCAGTTCGGGCGGAGAGATCAACCAAGAGATTCCTCCAGCAGAGGCTCGGA	3660
QY	1220	pSerAaA	1240
Db	3661	CAGCGCATCAATACCGGCTTCATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3720
QY	1240	AGUYyTh	1260
Db	3721	GGGACCATCAACAGAGATCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3780
QY	1260	uAaPProSerProLySGLyAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaA	1280
Db	3781	GGAAG	3840
QY	1280	RGUyUaA	1300
Db	3841	TGAGGTGATGATCTGTGACCCAGTCTCCCAAGAGAGAGAGAGAGAGAGAGAGAG	3900
QY	1300	oProAaA	1320
Db	3901	CCCCATAG	3960
QY	1320	gAa	1340
Db	3961	AGCCATCTCTCAGCAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCA	4020
QY	1340	aSerProAaA	1360
Db	4021	CAGCCCCACACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	4080
QY	1360	oAaRGSerTh	1380
Db	4081	TGATCTCTAGTGAAG	4140
QY	1380	GGUyUaA	1400
Db	4141	GAG	4200
QY	1400	nAa	1420
Db	4201	GAGGCTGAGGCGGCTGAAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	4260
QY	1420	uAa	1440
Db	4261	GAGGAGGCGGCTCATATCATAGATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG	4320
QY	1440	oLeuAaA	1460
Db	4321	CTTGAG	4380
QY	1460	pThGUYyAaA	1480
Db	4381	CACGAGGCGCTCAACAGCTGCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	4440
QY	1480	oGUYyAaA	1500
Db	4441	CGGCGGAG	4500
QY	1500	uAa	1520
Db	4501	ACGAGCTGCTTACAG	4560
QY	1520	ySerTh	1540

Dh 4561 CTCATTGGCGGCGGCGCCGGTCAATTGCTGAGCTGGTAAGCCGCGAGAGCCC 4620
Qy 1540 oLeuThrTYrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgIysSerProVa 1560
Db 4621 CCGACCTTAAGAAGACCAAGGGGACCCCTTGCCGGGCACTCCAGAGAGTTCCGCCGT 4680
Qy 1560 IThrMetAArgGluProThrProArgLeuGlnGlySerLeuSerSerIysAlaIse 1580
Db 4681 GACACAGCGGGAGACCCACGCGCGCTGAGAGGGGAGCGCTTTCGTCGACGAAGGATC 4740
Qy 1580 rGlnAspArgIysLeuThrSerThrProArgGluIleAlaIysSerProHisSerThrVa 1600
Db 4741 CAGAGACCGAAAGCTGAAGTGAAGCGCTCGTGAATGCGCAAGTCCCGAGACGACCGT 4800
Qy 1600 IProGlnHisHisProHisAsProIleSerProTYrGlnHisLeuLeuArgIyValSerGI 1620
Db 4801 GCCCGAGACACCAACCAACCCATCTCGCCCTAAGACACTGCTTCGGGGCGGTGAGTGG 4860
Qy 1620 yValAspLeuTYrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGI 1640
Db 4861 CGTGGACCTGTATCGAGCGACATCCCTGGCGCTTGACCCCACTCCATACCCGCGG 4920
Qy 1640 yIleProLeuAspAlaAlaAlaIATyTYrLeuProArgHisLeuAlaProAsnProTh 1660
Db 4921 CATCCCTCTGAGAGCGAGCGGCTGCTACTAGCTGCCCCGACACTGGCCCCCAACCCAC 4980
Qy 1660 rTYrProHisIleLeuTYrProProTYrLeuIleArgIyTYrProAspThrAlaIleuGI 1680
Db 4981 CTACCCGCGACCTGTACCCACCTTACCTCATCGCGGGTTACCCGACACGCGCGGCTGGA 5040
Qy 1680 uAsnAArgGlnThrIleIleAsnAspTYrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db 5041 GAACCGCGAGACCATCATCAATGATCAATCACCTCGACAGACATGACACCAACCGCGC 5100
Qy 1700 aThrAlaMetAlaGlnAlaArgAlaAspMetLeuArgIyLeuSerProArgGlnSerSerIe 1720
Db 5101 CACGCCATAGGCGCGAGAGCTGATATGCTGAAGGGGCTCTCGCCCCGAGAGTCTGGCT 5160
Qy 1720 uAlaLeuAsnTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisIle 1740
Db 5161 GGACCTCAACTACGCTGGCGGTCCCGAGGACATCATGACCTGTCCAGATGCGACACCT 5220
Qy 1740 uProValIleuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTY 1760
Db 5221 GCCTGTGCTCTGCTCCCGCGACACAGGACCCGACGACCCGATGACCGCTTCCCTA 5280
Qy 1760 rLeuProThrAlaProGlnProPheSerSerArgHisIserSerSerProIleuSerProGI 1780
Db 5281 CCTGCCACCGCGCGCCAGCTTCAGACAGCGCCACAGCAAGCTCCCACTTCCCAAGG 5340
Qy 1780 yGIYProThrHisLeuThrIysAspProThrThrThrSerSerSerGIuArgGIuArgAspAr 1800
Db 5341 AGGTCCAAACACTTGAACAAACCAACCAACCAAGTCTCTCGTCCGAGCGGAGGAGAACCG 5400
Qy 1800 gAspAArgGIuArgAspArgAspArgGIuArgGIuIysSerIleLeuThrSerThrThrTh 1820
Db 5401 GGATTCAGAGCGGAGACCGGATCGGAGCGGAGAAAGTCCATCTCACTGACCAACAGAC 5460
Qy 1820 rValGlnHisAlaProIleThrArgProGIyThrGlnGlnSerSerGIySerSerGIySe 1840
Db 5461 GGTGGAGCAGCGACCACTCGAGACCTGGTACAGAGCAG-----AGCAGCGGACAG 5511
Qy 1840 rSerGIyGIyGIyGIyGIySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5512 CAGCGCGGGGGGGGGGAGAGAGCGCGCGCTCCCACTCCCAATGCCACACAGCA 5571
Qy 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsa 1880
Db 5572 CTCGCCCATCTCCCTCGGACCCAGAGATGCCCTCCAGACAGACCAAGTGTGCTTCAAA 5631
Qy 1880 nThrGlnMetIysGlyIleIleThrAlaValGlnProSerIysProThrValLeuArgSe 1900
Db 5632 CACAGGACGTAAAGGATATCATCAACGCTGTGAGCGCCAGCAAGCCCAAGGTCTCTGAGGTC 5691

Qy 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisGysPr 1920
Db 5692 CACCTCACCTCTCCACCCGTTGCGCGGCTGGCAATTCACCTGACCAACCACTGCCCC 5751
Qy 1920 oLeuGIyGIyThrIleuAspGIyValTYrProThrIleMetGlnProValIleuIleuProIy 1940
Db 5752 ACTGGGCGGACCCCTGATGGGGGTCTACCTTACCTTCATAGAGCCCGCTGTGTGCCAA 5811
Qy 1940 sGlnAlaProArgValAlaArgProGIuArgProAlaArgAlaAspThrGIyHisAlaPheIle 1960
Db 5812 GAGAGCCCCCGGGGTGCCCCGGCAGAGCGCCCCAGACAGACCGGCGCATCCCTTCT 5871
Qy 1960 uAlaIysProProAlaIaArgSerGIyLeuGlnProAlaIserSerProSerIysGIySerGI 1980
Db 5872 CGCAGAGCCCCAGCCCGCTCGGGGTGAGCGCGCTCTCCGCCAGCAAGGAGCTCGGA 5931
Qy 1980 uProArgProLeuValProProValIserGIyHisAlaThrIleAlaArgThrProAlaIy 2000
Db 5932 GCCCGGCGCCCTAAGTGGCTCTGTCTGTGCAAGCGCACCATGCGGACCCCTTCGAA 5991
Qy 2000 sAsnLeuAlaProHisHisAlaIserProAspProProAlaProProAlaSerAlaSerAs 2020
Db 5992 GAACCTCGACCTCAGACCAAGCGGACCGCGGACCGCGGCGACCTGGCTCGGCTCGGA 6051
Qy 2020 pProHisArgGIuIysThrGlnIserIysProPheSerIleGlnIleuGlnIleuArgSe 2040
Db 6052 CCGGACACCGGAGAAAGACTCAAAAGTAAACCTTTTTCATTCAGGAAGTGAACCTCGGTC 6111
Qy 2040 rLeuGIyTYrHisGlySerSerTYrSerProGIyGIyValGlnProValIserProAlaIse 2060
Db 6112 TCTGGGTTTACACAGGAGAGCTACAGCCCCCAAGGGGTGAGCGGCTGTGTGAG 6171
Qy 2060 rSerProSerLeuThrHisAspIysGIyLeuProIyHisIleuGlnIleuAspIysSe 2080
Db 6172 CTCACCCACTCTGACCCCAAGCAAGGGGCTCCCAAGCACTGGAAGACTGACAAAGAG 6231
Qy 2080 rHisLeuGIyGIyIleuArgProIysGlnProGIyProValIysLeuGIyGIyGIuAl 2100
Db 6232 CCACTGAGAGGGAGCTCGGCGCCCAAGCAGCAGCGCCCGTGAAGCTTGCGCGGAGGCG 6291
Qy 2100 sAlaHisIleuProHisIleuArgProIleuProGIySerGlnProSerSerSerProIleuIe 2120
Db 6292 CGCCCACTTCCACACTGCGGCGCTGTGCTGAAGCCAGCTCTGCAAGCCGCTGCT 6351
Qy 2120 uGlnThrAlaProGIyValIysGIyHisGlnArgValAlaThrLeuAlaGlnHisIleSe 2140
Db 6352 CAGAGCGGCCCAAGGGGTCAAAAGTCAACAGCGGGTGTCACTTGCCCAAGCAACTAG 6411
Qy 2140 rGlnValIleThrGlnAspTYrThrArgHisHisProGlnGlnIleuSerAlaProLeuPr 2160
Db 6412 TGAAGTCATACACAGAGTACACCCGAGCACCAACCAAGACTGAGCGACCCCTGCC 6471
Qy 2160 sAlaProLeuTYrSerPheProGIyAlaIserCysProValIleuAspLeuArgArgProPr 2180
Db 6472 CGCCCCCTCTACTCTTCTTCTGAGGCGAGCTGCCCGGTCTGAGACTTCCGCGCCAC 6531
Qy 2180 oSerAspLeuTYrLeuProProProAspHisGlyAlaProAlaArgGIySerProHisIse 2200
Db 6532 CAGTGAACCTTACTCTCGGCCCCCGGACCAATGTGCGCGGCTCCGCCACAG 6591
Qy 2200 rGlnGIyGIyIysArgSerProGIuProAsnIysThrSerValIleuGIyGIyGIyIus 2220
Db 6592 CGAAGGGGAGCAAGGATCTCCAGAGCAAAAGAGTGGCTTGAGGTGTGTGAGGA 6651
Qy 2220 rGlnIleGIuProValIserProProGIyGIyMetThrGlnProGIyHisSerArgSerAl 2240
Db 6652 CGGATTTGAACCTGTGTCCCAACCGAGGGCAATGACGAGGACGAGGCACTCCCGAGTGC 6711
Qy 2240 aValTYrProLeuLeuTYrArgAspGIyGIyGlnThrGlnProSerArgMetGIySerIy 2260
Db 6712 TGTGTACCGGCTGTCTTACCGGGAATGGGAAACAGACGAGCCAGAGATGGGCTCCAA 6771

QY 2260 sSerProGlyAaAnthSerGlnProProAlaPhePheSerIyblLeuThrgIuSerAaSe 2280
DB 6772 GTCTCCAGGCAACACCGCCAGCGCTTCTTGAGCAAGCTGACCGAGCAATCTC 6831
QY 2280 FAlaMetValIySerIyblYgInGluIleAaAnthIyblLeuAnthrHisAaAnthGAs 2300
DB 6832 CGCCATGTCAGAGTCAGAGAGCAAGAGATCAACAGAGAGTGAACCCCAACCCGAA 6891
QY 2300 nGluProGluTyrAnthIleSerGlnProGluTyrGluIlePheAaMetProAlaIleTh 2320
DB 6892 TGAGCTGATACATATTCAGCCAGCTGGAGCGAGATCTTCAATATGCCCCGCAATAC 6951
QY 2320 rGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAaMe 2340
DB 6952 CGGAACAGGCTTATGACTATAGACGACGCGGTGAGGAACATGCGACAGCAACAT 7011
QY 2340 tGlyLeuGluAlaIleIleIleArgIyblAlaLeuMetGlyIyblYgInGluIlePhe 2360
DB 7012 GGGGCTGAGGCGCAATATGAAAGGACATCATGGGTAAATATACCAAGTGGAGAGTC 7071
QY 2360 rProProLeuSerAlaAaAnthPheAaProLeuAaAnthAlaSerIleuProAlaI 2380
DB 7072 CCGCGCGCTAGCGCCAAATGCTTTAACTCTGATATGCCAGTCCAGCTGCCCGCTGC 7131
QY 2380 aMetProIleThrAlaAlaAaPglYArgSerAaPHisThrLeuThrSerProGlyIyG 2400
DB 7132 TATGCCCATAAACGCTGCTGAGCGAGGATGACCAACATCACTCCGCGAGGTGGCGG 7191
QY 2400 yGlyIyblAlaIyblValSerGlyArgProSerSerArgIyblAlaIyblSerProAla 2420
DB 7192 CGGGAAGGCAAGGTCCTGCGAGAGCCAGAGAGCGAAGAGCCCAAGTCCCGCGCGCGG 7251
QY 2420 yLeuAlaSerGlyAaAaPglProProSerValSerSerValHisSerGluGlyAaPgl 2440
DB 7252 CTGGGACTCTGGGAGACCGGCCACTCTGTCTCTCTGACTCGAGGAGAGCTCAA 7311
QY 2440 nArgArgThrProLeuThrAaAnthArgValTyrGluAaPglProSerSerAlaGlySer 2460
DB 7312 CCGCGCGAGCGCGCTCAACAGCGCGTGTGGAGAGAGCGCTCCGCGAGGTTCCAC 7371
QY 2460 rProPheProTyrAaAnthProLeuIleMetArgIleuGlnAlaGlyAlaMetAlaSerPro 2480
DB 7372 GCCATTCCCTACCAACCCCTGATCATGCGGCTGCAAGCGGAGTGTCAATGCTCCACC 7431
QY 2480 oProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisIleAlaTyrAaPgl 2500
DB 7432 CCCACGGGCGCTCCCGGAGGAGCGGCGCTGCTGGCGCCGCCACACCGCTGGAGCA 7491
QY 2500 uGluProIyblProLeuLeuCySerGlnTyrGluThrLeuSerAaPgl 2517
DB 7492 GAGGCCAAGCACTGCTGCTGCTGCAAGTACAGAGACATCTCCAGACGAG 7543
RESULT 8
ACA62250 ID ACA62250 standard; cDNA; 7521 BP.
XX ACA62250;
AC 12-AUG-2003 (first entry)
XX
DT
XX
DE Human nuclear receptor corepressor SMRte coding region cDNA.
XX
XX Human; aa: gene; SMRte: nuclear receptor corepressor; gene therapy;
KM tissue typing; cancer.
XX
OS Homo sapiens.
XX
XX
FH Key location/Qualifiers
FT CDS 1..7524
FT /tag= a
FT /partial
FT /product= "SMRte"
FT /note= "No stop codon given"

XX US2003027137-A1.
PN
XX
XX 06-FEB-2003.
PD
XX
XX 27-MAR-2001; 2001US-00819104.
PF
XX
XX 29-MAR-2000; 2000US-0193138P.
PR
XX
XX (CHEN/) CHEN J D.
PA
XX
P1 Chen JD;
PI
XX
DR WPI: 2003-466139/44.
DR P-PSDB; ABU61812.
XX
PT New SMRte proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.
XX
PS Claim 2; Page 48-56; 90pp; English.
XX
XX The invention relates to an isolated SMRte nucleic acid molecule. The
CC nucleic acids are useful in gene therapy, as hybridisation probes for
CC identifying SMRte-encoding nucleic acid molecules and as primers for
CC amplifying of SMRte nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRte are useful for treating or
CC preventing a condition associated with aberrant SMRte protein or nucleic
CC acid expression or activity, such as cancer. The present sequence
CC represents the human nuclear receptor corepressor SMRte coding region
CC cDNA
XX
SQ Sequence 7521 BP; 1635 A; 2728 C; 2212 G; 946 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 7521
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3
US-09-522-753-5 (1-2517) x ACA62250 (1-7521)
QY 1 MetSerGlySerThrgInleuValAlaGlnThrTyrArgAlaThrgIuProArgTyrPro 20
DB 1 ATGTGGGCTCCACAGAGCTGTGGACAGAGCTGAGGAGGCACTGAGCCCCGCTACCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAaPglIyLeuLeu 40
DB 61 CCCACAGGCTTCTTACCAAGAGATGAGATGCCCGAGCGGACAGCGAGTGGGCTCTCG 120
QY 41 GluTyrGlnHisIleSerArgAaPglYrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACAGACACACCGCGGACTAGCTGCTCCACCTGCGCGGCTCCATCATTCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAaAnthGlySerGln 80
DB 181 CCCACGGGCGGAGGCTCTCCCTGCTGCTGAGTTCAAGCCCGGAAATGAACGCTCCAG 240
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluIleuGlyIyblYgInMet 100
DB 241 GAGCTCACCTGGGCGGAGATGCCACTATACCTGCCGAGGTGGGAAAGTCAAGATG 300
QY 101 GluPheIleGluSerIyblYgInProAlaGluGluLeuLeuProAaPglProLeuAaPglPro 120
DB 301 GAGTTCATTGAAAGCAAGGCGCTCGGCTAGAGCTGCTGAGACCCCTGCTGCAACG 360

QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleYsaAspArgSer 140
DB 361 TCACCCCTGCTGGCCACGGGCCAGCTGCGGGATCTGAAGAATCTCAACAAAGACCGTAGC 420
QY 141 LeuThrGlyIleYsaLeuGluProValSerProProSerProProHieThrAspProGluLeu 160
DB 421 CTGACGGGCAAGCTGGAAACCGGTGTCTCCCGACGCCCCCGACACTGACCTTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerIleYsaGlnIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGCTCCCGCCACGCTGCTCCAAAGAGAGACTATCCAGAAACATGGACCCCGTGGAC 540
QY 181 ArgGluIleThrMetValGlnGlnGlnIleSerIleYsaIleYsaIleYsaGlnGlnIleu 200
DB 541 CAGAGATACCACTGATAGCAGCAGATCTTAAGCTGAAGAAAGACGACCAACAGCTTG 600
QY 201 GlnGlnGlnAlaAlaIleYsaProProGluProGluIleYsaProValSerProProIleGlu 220
DB 601 GAGGAGAGAGCTGCGCAAGCGCGCGAGCGCTGAGAAACCGGTGTACCGCGCGCCATCGAG 660
QY 221 SerIleYsaIleArgSerLeuValGlnIleIleIleYsaAspGluAsnArgIleYsaIleAla 240
DB 661 TCGAAGCACCGCAGCTGCTGCGAGATCATCTACAGAGAAACCGGAAGAGCTGAAGCT 720
QY 241 AlaIleIleArgIleIleuGlnGlnGlnGlnIleProGlnIleValGlnLeuProLeuThrAsnGlnPro 260
DB 721 GCACATCGATCTGGAAGCGCTGCGGCGCGCGAGCTGAGCTGCTGTATCAACACGAGCCC 780
QY 261 SerAspThrArgGlnIleYsaIleGlnIleYsaIleIleYsaIleIleYsaIleIleYsaIle 280
DB 781 TCCGAGACCGCGAGATCATGAGAAATCAAAATTAACCGGCGATGCGGAAGAAAGCTA 840
QY 281 IleLeuYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 300
DB 841 ATCTTCTACTTCAAGAGAGAGAAATCAACCTCGGAAACATGGAGAGAAAGTTCTGCCAG 900
QY 301 ArgIleYsaAspGlnIleuMetGlnIleuGlnIleYsaIleGlnIleGlnIleuAsnPro 320
DB 901 CCTATGACCAAGCTCATGAGCGCTGCGGAAGAAAGTGGAGCGCATCTCGAACAACCCC 960
QY 321 ArgArgArgAlaIleuGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 340
DB 961 CCGCGCGCGCGCAAGAGAGCAAGGTTCGAGATCTACAGAAAGCAATGTTCCCTGAGATC 1020
QY 341 ArgIleYsaIleYsaIleuGlnIleuGlnIleuArgMetGlnIleYsaIleGlnIleYsaIle 360
DB 1021 CCGAAGCACCGCAGCTGCGAGAGAGCCCAATGCAAG--AGGTTGGCCAGCGCGGCAATGG 1077
QY 361 LeuSerMetSerAlaAlaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 380
DB 1078 CTGTCCATGTCCGCCCGCGCGAGCGAGCAAGAGTGTCAAGAGATCATTCATGCGCTCTCA 1137
QY 381 GlnGlnGlnIleuAsnIleuGlnIleuYsaIleuMetArgGlnIleuAlaIleProProMetLeuYsa 400
DB 1138 GAGCAGAGAAATCTGGAAGAGAGATGCGCAGCTGCGGTGTCCCGCCCAATGCTGTAC 1197
QY 401 AspAlaAspGlnIleuArgIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 420
DB 1198 GAGCGTCAACAGCAGCGCATCAAGTTCATCAACATGAACGGCTTATGGCGGACCCCAATG 1257
QY 421 IYsaValIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 440
DB 1258 AAGGTGTACAAGACCGCGAGTCAATGAACATGTGAGTGAAGAGAGAAAGAGAGACTTTC 1317
QY 441 ArgGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 460
DB 1318 CCGAGAAAGCTCATGAGCATCCCAAGAACTTGGCTGATGCTCATTCCTCGAGAGAG 1377
QY 461 IYsaThrValAlaGlnCysValIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 480
DB 1378 AAGACAGTGGCTGAGTGGCTCTCTATTAATCACTGAAGAAAGATGAAGAACTTAAG 1437
QY 481 SerLeuValAlaArgSerIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 500

DB 1438 ACCCTGTGAGACGAGAGCTATGCGCCCGGCAAGACACACAGCAAAACAGCGCAG 1497
QY 501 Gln 520
DB 1498 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
QY 521 AspGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 540
DB 1558 GATGAGAAAGAGAAAGAAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
QY 541 AspIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 560
DB 1618 GACAAAGAAAG 1677
QY 561 IYsaGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 580
DB 1678 AAGAGAGCTGTGGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
QY 581 ArgIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 600
DB 1738 CCGATCAACCGCTCAATGAGCTAATGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
QY 601 SerAlaGlnIleuAlaSerMetGlnIleuAsnGlnIleYsaIleYsaIleYsaIleYsa 620
DB 1798 AGCGCGAGCTGCGCTCCATGAGCTGAATGAGATTCTGCTGACAGAAAGAAATG 1857
QY 621 GlnThrAlaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 640
DB 1858 GAAACAGCCAAAGAAAGTCTCTGGAACACGCGCGCAATGCTGCGCATTCGCCGCGAG 1917
QY 641 ValIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 660
DB 1918 GTGGGCTCCAAGACTGTGCGAGGTGAAGAACTTCACTTAACATCAAGAAAGAGCAG 1977
QY 661 AsnIleuAspGlnIleuGlnGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 680
DB 1978 AACCTCGATGAATCTTGAAGAGCAAGCTGAAGATGAAGAGAGAGAGAGAGAGAG 2037
QY 681 ArgIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 700
DB 2038 AGGAG 2097
QY 701 GlnAspGlnIleuMetGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 720
DB 2098 GAGGATGAGAGATGAGAGCGCTGCGGCGTGAAGCGGAAATGAAGAGAGATGTTGAGAG 2157
QY 721 AlaGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 740
DB 2158 GCTGAA-----GCC 2166
QY 741 ThrValAsnAsnSerSerAspThrGlnIleProSerProHieThrGlnIleAlaIleYsa 760
DB 2167 ACTGTCAACAACAGCTCAACACAGAGACATCCCTCTCTCAACATGAAGCGCGCAAG 2226
QY 761 AspThrGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 780
DB 2227 GACACAGGCGCAAAATGGCGCAAGCCCGAGCAACCTGTGGCGCGAGCGGCGCACCCCA 2286
QY 781 GlnProProThrProProArgArgThrSerArgAlaProIleGlnProThrProAlaIle 800
DB 2287 GGGCCACCCACCCACACAGAGAGAGATCCGGCC-CCCACTGAGTCCACCCCGGCTTC 2345
QY 800 GlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIle 820
DB 2346 TGAAGCACTTAAGCCCTCAAGCCCGCCACAGACACCCCAATTCCTCTTCACTCTCTCC 2405
QY 820 OValValProYsaGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 840
DB 2406 TGTGGTCCCAAG 2465
QY 840 YGlnGlnIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsaIleYsa 860

Dh 2466 GGAGAGCAGAAAGCCCCCGGCTGAGAGAGCTGGACGTGACACAGAGGAAGCCGAGGA 2525
Qy 860 uProValIySeSerGIuCyThrThrgIuGluAgluGluGlyProAlaIySgIyLysAaPAI 880
Db 2526 GCCCGTCAAGAGCAGAGTGCACGAGAGAGCCGAGAGAGGAGGCGGCGCAAGGAGAGAGC 2585
Qy 880 agIuaIaIaGluAaThraIaGluGlyAlaLeuIySgIyLysAaPAIuIySgIyGlySe 900
Db 2586 GGAGAGCGCTGAGAGCAGAGGCGGAGAGGCGCTCAAGGAGAGAGAGAGAGAGGAGGAG 2645
Qy 900 rGIyArGIaIaThrThraIaIySeSerGIyAlaProGluIaAspSerIaAspSerIaTh 920
Db 2646 CGGAGAGGAGCAGCAGCAGCAGAGAGCTCGGAGGCGGCGGAGAGAGAGAGAGAGAG 2705
Qy 920 rCySeSerAlaAspGIuValIaAspGIuAgluGluGlyAlaAspLysAaPAIuIySgIySe 940
Db 2706 CTGACAGTGCAGAGAGAGTGCAGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAG 2765
Qy 940 oArGIProSeSerIeuthrThrThrgIyAspProArGIaAspIaAspIaSeSerProGluIy 960
Db 2766 AAGGCGCAGGCTCTTCACCCGAGCTGGAGAGCCCGGAGAGAGAGAGAGAGAGAGAG 2825
Qy 960 oIeuaAspLeuIySgIuIeulYsGluAaIaIaIaIeProProIleGIuAaIaThrIy 980
Db 2826 ACTGGAAGCTGAAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2885
Qy 980 sValIaIaGluProProArGIuAaPAIaIaProThIyLysAaPAIaProProIaAProPr 1000
Db 2886 AGTCATAG 2945
Qy 1000 oProProGIaAspLeuGIuIeulYsGluAaIaIaIaIeProProIleGIuAaIaThrIy 1020
Db 2946 GCGAGCGCAAAAGCTGAG 3005
Qy 1020 gGIyLysSeSerIySeSerProAlaProProAlaAspLysGlu----- 1033
Db 3006 GGGCAG 3065
Qy 1034 -----AlaPheAlaIaGluAaIaGluIyLysAaPAIaAspProProCyTrpThrIySe 1052
Db 3066 CCGAGAGCTTGCAG 3125
Qy 1052 yLeuProPheProValProProArGIuAaIaIeYsAaIaSeSerProIaIaAProAaPr 1072
Db 3126 CTGAGCGCTTCCCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3185
Qy 1072 oSerAlaIaPheSerIyAlaProProGIyIaIaProLeuProLeuGIyLeuIaIaAspThraI 1092
Db 3186 CTGAGCGCTTCCCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3245
Qy 1092 aArGIProValIeProArGIProProThrIleSeSerAaProProProLeuIleSeSerAl 1112
Db 3246 CCGAGCGCTTCCCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3305
Qy 1112 aIySgIaIaProSeSerValIeulYsGluAaIaIaIaIeSeSerGIyMeSeSerValGI 1132
Db 3306 CAG 3365
Qy 1132 nIeulIa 1152
Db 3366 GCTGCGAG 3425
Qy 1152 uProLeuProMeArGIProIyLysAaPAIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1172
Db 3426 GCGCGCTTCCCGTGCAG 3485
Qy 1172 uSeSerProArGIyGluAaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1192
Db 3486 GTCGCGAG 3545
Qy 1192 aSeSerValIeulYsGluIyThraIaIeulYsSerValProGIyGlySeSerIeThrIySgIy 1212
Db 3546 GTCCGCTGAG 3605

Qy 1212 eProSeThrArGIaIaProSeArSeSerAlaIeThrThrArGIySeSerIeThrIaIaIa 1232
Db 3606 TCCAG 3665
Qy 1232 yThrProAlaAspValIeulYySgIyThrIleThrArGIleIleGIyGluAaSeSerPr 1252
Db 3666 CAGCGAG 3725
Qy 1252 oSerArGIuAaAspArGIyArGIuAaAspSerIeProIySgIyIaIaIaIeThrIyGlu 1272
Db 3726 GAGTGGCTTGCAG 3785
Qy 1272 yLysIySgIyIa 1292
Db 3786 CAG 3845
Qy 1292 uAspGIyArGISeSerSeSerGIyProProIaIaIaIaIaIaIaIaIaIaIaIaIa 1312
Db 3846 GAG 3905
Qy 1312 pMeuMeGIuGIyArGIa 1332
Db 3906 CATGATGAG 3965
Qy 1332 yArGIa 1352
Db 3966 CCGTGCATTCGCGGAG 4025
Qy 1352 gGIySeSerIeThrGIuGIyIleProArGISeSerIyValGIuAaIaIaIaIaIaIaIa 1372
Db 4026 CGGCTCATACAG 4085
Qy 1372 gArGIuAaIa 1392
Db 4086 TCGAG 4145
Qy 1392 pIeulThrGIuAaIa 1412
Db 4146 CCGAG 4205
Qy 1412 uGIyLeuValaIa 1432
Db 4206 GGGCTGTGTGAG 4265
Qy 1432 uIeuaArGIa 1452
Db 4266 GCTGCGAG 4325
Qy 1452 rGIuGIyThrProLeuIySeSerIyArGIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1472
Db 4326 GAG 4385
Qy 1472 pValaIa 1492
Db 4386 GGTAGCTTCCCGTGCAG 4445
Qy 1492 IMeAlaAspAlaIa 1512
Db 4446 GATGCGAG 4505
Qy 1512 oGIyThraIa 1532
Db 4506 AGGAG 4565
Qy 1532 uIeulYyProArGIuIa 1552
Db 4566 GCTGGGTAAAG 4625
Qy 1552 yIa 1572
Db 4626 CCACTTCCAG 4685

QY 1572 YSerLeuSerSerSerIyAlaSerGIuAspArgIyLeuThrSerThProArgGIu11 1592
| | | | |
Db 4686 CAGCCTTTGTCAGAGAACGATCCAGAGACGAAAGCTGACGTCAGCCCTGTCAGAT 4745
| | | | |
QY 1592 eAlAlaYSerProHISerThValProGIuHIShISpProHISProI1SerProIyrg1 1612
| | | | |
Db 4746 CGCCAAAGTCCCGCAGACGCGTGCAGACCAACCCACATCTCGCCCTATGA 4805
| | | | |
QY 1612 uHISLeuLeuArgGIyValISerGIyValAspLeuTYrArgSerHISelProLeuAlaH 1632
| | | | |
Db 4806 GCACCTGCTTCGGGGCGTAGAGTGCCTGGAACCTGTATCGCAGCAACATCCCTGGCCCTT 4865
| | | | |
QY 1632 eAspProThSerI1eProArgGIy11eProLeuAspAlaAlaAlaIyTYrTYrLeuPr 1652
| | | | |
Db 4866 CCACCCCACTCAATACCCCGCGGCATCCCTTGAGCAGCAGCGGCTGCTACTACCTGCGC 4925
| | | | |
QY 1652 oArghISleuAlaProAsnProThTYrTYrProHISleuTYrProProTYrLeu11eArg1 1672
| | | | |
Db 4926 CCGACACCTGGGCCCCCAACCCCACTACCCGCACTGTACCCCACTTACTCTACCTCGCGG 4985
| | | | |
QY 1672 YTYrProAspThraAlaAlaLeuGIuAsnArgInThrI1e1eAsnAspTYr11eThrSe 1692
| | | | |
Db 4986 CTAACCCGACAGCGCGCGCTGGAAGAACGGCAGACATCATATGACTACATCACTC 5045
| | | | |
QY 1692 rGIuInMeThSHISAsnThra1aThra1aMeTa1aGIuArg1aAspMe1euaArg1 1712
| | | | |
Db 5046 GCAGCAGATGACACCAACACCGGCACCGCATGGCCAGCGAGCTGATATGCTGAGGGG 5105
| | | | |
QY 1712 YLeuSerProArgIuISerSerIeua1aLeuAsnTYrAlaAlaGIyProArg1y11e11 1732
| | | | |
Db 5106 CCTCTGCCCCGGAGCTCTGCTGAGCTCACTCACTACGCTGCGGGTCCCCGAGGCATCAT 5165
| | | | |
QY 1732 eAspLeuSerGIuValProHISleuProValLeuVal1ProProThProGIyThrProAl 1752
| | | | |
Db 5166 CCACTGCTCCCAAGTGCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5225
| | | | |
QY 1752 aThra1aMeTa1aAspArgLeuAlaTYrLeuProThra1aProGIuProHISerSerArgHIS 1772
| | | | |
Db 5226 CACCGCATGAGACGCGCTTGCTTACCTCCCAACCGGCCCCAGCCCTTGAGCAGCGCCA 5285
| | | | |
QY 1772 sSerSerSerProLeuSerProGIyGIyProThrHISleuThTyLeuProThrThrHISe 1792
| | | | |
Db 5286 CAGCAGCTCCCACTCTCCCAAGAGGTCCCAACACTTGACAAACCAACCAACCAACCTGTC 5345
| | | | |
QY 1792 rSerSerGIuArgGIuArgAspArgAspArgGIuArgAspArgAspArgGIuArgGIu1y 1812
| | | | |
Db 5346 CTCGTCCGAGCGGAGAGACCGGAGTCGAGAGCGGAGCCGGATCGGAGCGGAGGAAAA 5405
| | | | |
QY 1812 sSerI1eLeuThrSerThThThThValGIuHISAlaProI1eTrpArgProGIyThrg1 1832
| | | | |
Db 5406 GTCCATCTCTCACGTCACCAACGAGGTGAGACGCAACCCATCTGGAGACTGTGTACAGA 5465
| | | | |
QY 1832 uGIuISerSerGIySerSerGIySerSerGIyGIyGIyGIyGIySerSerSerArgProAl 1852
| | | | |
Db 5466 GCAAGACACGCGGAGAGCGGAGCGGAGCGGCGGGGGGAGCAGCAGCAGCGCGCCCGC 5525
| | | | |
QY 1852 aSerHISerHISAlaHISgluHISserProI1eSerProArgThGIuAspAlaLeuGI 1872
| | | | |
Db 5526 CTCCTCACTCCCATGCCCAACGACATCTGCCATCTCCCTCGCAGACCAAGATGCCCTCCA 5585
| | | | |
QY 1872 nGIuArgProSerValLeuHISAsnThrgIyMeTyGIy11e11eThra1aValGIuPr 1892
| | | | |
Db 5586 GCAGAGACCCAGAGTGTCTTCAACAACAGGCAATGAAGGATATCATCACCCCTGTGGAGCC 5645
| | | | |
QY 1892 oSerTySerProThValLeuArgSerThSerThSerSerProValArgProAlaAlaTh 1912
| | | | |
Db 5646 CAGCAAGACCCACGCTCTGAGGTCCACCTTCACCTCTCCACCCGTCGCCACCTGCGCAC 5705
| | | | |
QY 1912 rPheProProAlaThHIScYspProLeuGIyGIyThrLeuAspGIyValTYrProThrLe 1932
| | | | |
Db 5706 ATTCCCACTGCAACCACTGCGCACTGGGCGGACCTCGATGGGGTCTAACCTTACCT 5765
| | | | |
QY 1932 uMetGIuProValLeuLeuProIySGIuAlaProArgValAlaArgProGIuArgProAr 1952
| | | | |

Db 5766 CATGAGCCCGCTTGCTGCCCCAAGAGGCCCGCGGGTGCCTCCGCGCAGACGGCGCCG 5825
| | | | |
QY 1952 gAl1aAspThrgIyHISAlaPheLeuAlaTyProProAlaArgSerGIyLeuGIuProAl 1972
| | | | |
Db 5826 AGCAGACACCGGCATGCTCTTCTCGCAGAGCCCCAGCCCGCTCGGGCTGAGGCCGC 5885
| | | | |
QY 1972 aSerSerProSerTyGIySerGIuProArgProLeuVal1ProProValISerGIyHISAl 1992
| | | | |
Db 5886 CTCCTCCCCCAAGAGGCTCGAGGCCCGGCGCCCTAGGCTCTGCTGTCGCGCACGC 5945
| | | | |
QY 1992 aThrI1eAlaArgThrProAlaTyAsnLeuAlaProHIShISAlaSerProAspProBr 2012
| | | | |
Db 5946 CACCAATGCCCCACCCCGCCAGAAACCTGCACCTTCAACAGCCGACCGGACCGGC 6005
| | | | |
QY 2012 oAlaProProAlaSerAlaSerAspProHISArgGIuTySerThGIuISerTySerProHISe 2032
| | | | |
Db 6006 GCGCGCACTGCTCGCGCTCGGCTCGGACCCGACCGGAGAAAGACTCAAAAGTAAACCTTTTC 6065
| | | | |
QY 2032 r11eGIuInGIuLeuGIuLeuArgSerIeugIyTYrHISGIySerSerTYrSerProGIuGI 2052
| | | | |
Db 6066 CATCCAGGAACGTGAACCTCGTTCCTGAGTTACACGCGCAGACGCTTACAGCCCCGAAAG 6125
| | | | |
QY 2052 YValGIuProValISerProValISerSerProSerLeuThrHISAspTyGIyLeuProIy 2072
| | | | |
Db 6126 GGTGAGCCCGTCAAGCCCTGTAGCTCACAGCTTGACCCAGACGAAAGGCTCCCA 6185
| | | | |
QY 2072 SHISleuGIuGIuLeuAspTySerHISleuGIuGIyLeuLeuArgProIySGInProGI 2092
| | | | |
Db 6186 GCACCTGGAAGACTGAGAACAGACCACTGAGGGGAGCTGCGGCCAAGAGCCAGCAG 6245
| | | | |
QY 2092 YProValTyLeuGIyGIyGIuAlaAlaHISleuProHISleuArgProLeuProIySe 2112
| | | | |
Db 6246 CCCCCTGAAGCTTGGGGGAGGCGCCCACTCCACACTCGCGCCCTGCTGTAGAG 6305
| | | | |
QY 2112 rGIuInProSerSerSerProLeuLeuGIuThra1aProGIyValTyGIyHISgluArgTy 2132
| | | | |
Db 6306 CCAGCCCTGTCAGGCCCTGCTCTCCAGACCGCCCAAGAGGTCAAAAGTCAACAGCGGGT 6365
| | | | |
QY 2132 IValThrLeuAlaGIuHISelSerGIuVal11eThrgInAspTYrThArgHIShISPr 2152
| | | | |
Db 6366 GGTCACTCGGCCAGCAATCATGAGGTATCATACAGAGACTACACCGGACCAACCC 6425
| | | | |
QY 2152 oGIuInLeuSerAlaProLeuProAlaProLeuTYrSerPheProGIyAlaSerCySP 2172
| | | | |
Db 6426 ACAGCAGCTCAAGGACCCCTGCCCGCCCTCACTCTCTCCCTGGGGGCCAGCTGCC 6485
| | | | |
QY 2172 oValLeuAspLeuArgArgProProSerAspLeuTYrLeuProProProAspHISgluAl 2192
| | | | |
Db 6486 CGTCTGGAACCTCCGCGGCCCAACCAGTACCTTACCTCCGCGCCCGGACCATGTGC 6545
| | | | |
QY 2192 aProAlaArgGIySerProHISerGIuGIyGIyTyAsArgSerProGIuProAsnTyTh 2212
| | | | |
Db 6546 CCGGCGCCGTGCTCCCGCCCAACGAGGAGGAGGAGAGAGTCTCAGAGCAAAACAAGAC 6605
| | | | |
QY 2212 rSerValLeuGIyGIyGIyGIuAspGIy11eGIuProValISerProProGIuGIyMeTh 2232
| | | | |
Db 6606 GTCGGTCTTGGGGTGTGAGGAGCGGTATTGAACCTGTGTCCCAACCGGAGGGCATGAGC 6665
| | | | |
QY 2232 rGIuProGIyHISerArgSerAlaValTYrProLeuLeuTYrArgAspGIyGIuGIuH 2252
| | | | |
Db 6666 GAGGCCAGGACATCCCGAGGTGTGTATCCGCTGCTGTACCGGAGTGGGAGAACAGAC 6725
| | | | |
QY 2252 rGIuProSerArgMeGIySerTySerProGIyAsnThrSerGIuProProAlaPhePh 2272
| | | | |
Db 6726 GAGGCCAGCAGATGGCTTCAAGTCTCAGGCAACAACAGCAACCGGCAACCTTCTT 6785
| | | | |
QY 2272 eSerTyLeuThrGIuISerAsnSerAlaMeTyValTySerTyTyGIyGIuGIuLeuTy 2292
| | | | |
Db 6786 CAGCAAGCTGACCGAAGACCACTCCGCAATGTGTCAAGTCCAGAAAGCAAGAGATCAACAA 6845
| | | | |
QY 2292 sTyLeuAsnThrHISAsnArgAsnGIuProGIuTYrZAsnI1eSerGIuProGIyThrg1 2312
| | | | |

Db 6846 GAAGCTGACACCCGACCAACCGGAATGAGCCTGAAATACATATACGACGCTGGAGCGA 6905
Qy 2312 uilePheAaMePProAlaIleThrGlyLeuMeThrTyraArgSerGlnAlaVa 2332
Db 6906 GATTTTCAATATGCCCCCATACCGGAACAGGCTTATGACCTTATGAAAGCCAGGCGGT 6965
Qy 2332 IginglunHlaAaSerThraMetGlyLeuGlnAlaIleIleArgLysAlaLeuMetG 2352
Db 6966 GCAGGAACATGCCGACCAACATGGGGCTGGAGCCATATTTGAAAGGACCTCATGGG 7025
Qy 2352 ylyeTyraSerGlnTTPgluIuSerProProLeuSerAlaAaAlaPheAaPProLeuAs 2372
Db 7026 TAAATATGACCAAGGAGAGGTCCCGCGCTCAGGCGCCATCTTTTAACTCTCGAA 7085
Qy 2372 naLaSerAlaSerIuProAlaAlaMetProIleThraAlaAlaAaPglyArgSerAaPH 2392
Db 7086 TGCCAGTGCAGCCTGCCCGCTGTATGCCCCATTAACGCTGTACGAGACGAGTGAACA 7145
Qy 2392 sThrLeuThraSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
Db 7146 CACACTCACTCGCCAGGTGGCGGGAGGAGCCAGAGTCTTGAGACCCAGACCGG 7205
Qy 2412 glyValAlaLysSerProAlaProGlyLysAlaSerGlyAaAaArgProProSerValSerSe 2432
Db 7206 AAAAGCCMAGTCCCGGCGCCCTGGCATCTGGGGACCGGACACCTCTGTCTCTC 7265
Qy 2432 rValHlaSerGlnGlyAaPcyAaAaArgArgThraProLeuThraAaArgValTTPgluAs 2452
Db 7266 AGTGCACTCGAAGGAGACATGCAACCGCCGACCCCTCAACCAACCGCGTGGAGAGA 7325
Qy 2452 PaTgProSerSerAlaGlySerThraProPheProTyraAaPProLeuIleMeArGLeuG 2472
Db 7326 CAGGCGCTCGTCCGAGAGTTCAGAGCATTCCTTACAACCCCTGATCATGGGCTGCA 7385
Qy 2472 naLaGlyValMeArLaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7386 GGGGGGTGTATGGCTTCCCAACCCCAACCGGGCTCCCGGGCGAGCGGGCCCTCGC 7445
Qy 2492 agLyProHlaHlaIleThraAaPgluIuProLysPProLeuLeuCySerGlnTyrrGluTh 2512
Db 7446 TGCCCCCACCACAGCTGGGAGCAAGAGCCCAAGCCACTGCTGTGCAATACGAGAC 7505
Qy 2512 rLeuSerAaPSerGlu 2517
Db 7506 ACTCTCGACAGAGGAG 7521

RESULT 9
ADL13812
ID ADL13812 standard; DNA: 7524 BP.
XX
AC ADL13812;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #344.
XX
KM da: gene: osteopathic; antiinflammatory; antiarthritic; gene therapy;
KM joint space narrowing; osteophyte development; joint pain;
KM osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN MO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002MO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schaffer A;

XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 344; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 7524 BP; 1636 A; 2728 C; 2213 G; 947 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 7524
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3
US-09-522-753-5 (1-2517) x ADL13812 (1-7524)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThraArgAlaThrgluProArgTyrrPro 20
Db 1 ATGTGGGCTCCACACAGCTGTGGACACAGCTGAGGGCCACTGAGCCCGCTACCGG 60
Qy 21 ProHlaSerLeuSerTyrrProValGlnIleAlaArgThraThraAaPValGlyLeuLeu 40
Db 61 CCCACAGCTTCTTCCATCCAGTGCAGATGCCCCGAGACGACAGCATCGGGCTCTCG 120
Qy 41 GluTyrrGlnHlaHlaSerArgAaPtyrrAlaSerHlaLeuSerProGlySerIleIleGln 60
Db 121 GAGTACACAGACACACTCCGCGACTATGCTCCACCTGCGCCGCTCATCATCCAG 180
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGlnPheGlnProGlyLysGlnuArgSerGln 80
Db 181 CCCACGGGGAGGCGCTTCCCTGCTGTGAGTTCAGAGCCCGGAATGAACGATCCAG 240
Qy 81 GluLeuHlaLeuArgProGluSerHlaSerTyrrLeuProGluLeuGlyLysSerGlnMet 100
Db 241 GAGCTCACCTGGGGCCAGAGTCCACTATCACTCCGAGCTGGGGAAGTCAAGATG 300
Qy 101 GluPheIleGluSerLysArgProArgLeuGlnLeuLeuProAaPProLeuLeuArgPro 120
Db 301 GAGTTCATTGAAGAAGCAGCGCTCGCTAGACTGCTGCTGACCCCTGCTGGCCAGC 360
Qy 121 SerProLeuLeuAlaThrglyGlnProAlaGlySerGlnuAaPLeuThraAaPArgSer 140
Db 361 TCACCCCTGCTGCGCACGGGCGAGCTGGGGATGTGAAGACTCAACGAAGACCGAGC 420
Qy 141 LeuThrglyLysLeuGluProValSerProProSerProProHlaThraAaPProGluLeu 160
Db 421 CTGACGGGAGAGTGAACCGGTGTCTCCCGCAGGCCCGCCGACACTGACCTTGAGCTG 480
Qy 161 GluLeuValProProArgLeuSerLysGlnGluLeuIleGlnAaMetAaPArgValAaP 180
Db 481 GAGCTGTGCTCGCCACAGGCTGTCTCAAGAGAGACTGATCCAGAACATGACCGCGTGAC 540

OY	181	ArgGluIlethrmVetValGlueGlnGlnIleSerIysLeuIleIysIleGlnGlnIleu	200
Db	541	CGAGGAATCACCATGATGAGACAGAGATCTCTAAGCTGAAGAAGAACACCAACAGCTG	600
OY	201	GlueGlnGluIleAlaIlyProProGluIuProGluIySProValSerProProIleGlu	220
Db	601	GAGAGAGAGGCTGCCAACCCGCCAGGCTGAGAAAGCCCGTGTACACGCCGCCATCGAG	660
OY	221	SerIyshiSaRgSerIeuValGlnIleIleTyRAspGluAmaRgIySlySAlaGluAla	240
Db	661	TGGAAGACCCGACGCTGTGCAGATCATCTTACGACGAAACCGGAGAAAGCTGAAGCT	720
OY	241	AlaHisaRgIleIeuGlnGlyIleuGlyProGlnValGluIeuProIeuTyRAsnGlnPro	260
Db	721	GCACATCGAATTCGGAAGGCTCGGGGCCCAAGGTGAGCTGGCGCTGTACACCAAGCC	780
OY	261	SerAspTTrArGlnTyRhIsgIuAsnIleIySileAsnGlnAlaIameArgIySleu	280
Db	781	TCCGACACCCGGCAGTATCATGAGAAACATCAAAATAAACAGGCGATCCGGAAGAAGCTA	840
OY	281	IleIeuTyRpheIySArGArgAsnHisAlaRgIySIntTyPlySgInIySpheCysGln	300
Db	841	ATCTTGATCTTCAAGAGAGGAAATCACCTCGGAAACAATGGAGACAAATTTCTGCAG	900
OY	301	ArgTyRAspGlnIleuMetGluAlaIeuGlnIySlySlyValGluIuRgIleIuAsnAsnPro	320
Db	901	CGTATATGACCAAGCTCATGAGGCTCGGAGAAAGAGTGGAGCGCATGAAGAACACCC	960
OY	321	ArgArGArGAlaIySgluSerIySAlaArgGlnTyTyRgIuIySgInPheProGluIle	340
Db	961	CGGCGGCGGGCCAGAGAGACAGAGCTTGCAGATCATACGAAGACATTTCCGTGAGTC	1020
OY	341	ArgIySgInaRgIuIeuGlnGlnIuArgMetGlnSerArGValGlyGlnaRgIySergIy	360
Db	1021	CGCAAGCAGCGGAGACTGCAGAGCGCATGCGAG--AAGGTGGCGACGCGGCGAGTGGG	1077
OY	361	IeuSerMetSerAlaIaIaRgSerGluHisGluValSerGluIleIeAspGlyIeuSer	380
Db	1078	CTGTCCATGTGCCCCGCCGACGACGACGACGAGGTGTCAGATATGATGAGCTCTCA	1137
OY	381	GluGlnGlnuAsnIeuGlnIySgInMetArgGlnIeuAlaValIleProProMetIeuTyR	400
Db	1138	GAGCAGAGAGAACTGTGAGAGAGCATGGCGCAGCTGGCGGTGATCCGCCCATGTGTAC	1197
OY	401	AspAlaAspGlnGlnaRgIleIySpheIleAsnMetAsnGlyIeuMetAlaAspProMet	420
Db	1198	GACGCTGACCAAGCAGCAGCATCAAGTTCTCAACAAGAACGGGCTTATGCCGACCCCATG	1257
OY	421	IySValTyTyRySAspArgGlnValMetAsnMetTPSerGlnGlnIuIySgIuThRphe	440
Db	1258	AAAGGTGTCAAAGACCGGACAGGTATGAAACATGTGAGTGAACAGAGAAAGGACCTTC	1317
OY	441	ArgGluIySPhemeGlnHisProIySAsnPheGlyIeuIleAlaSerPheIeuGluIuRg	460
Db	1318	CGGAGAAAGTTCAATCAGCATCCCAAGAACTTTGGCGATCGCATCTTCTGTGAGAGG	1377
OY	461	IySThrValAlaGluCysValIeuTyTyTyRLeuThrIySlySAsnGluAsnTyRlys	480
Db	1378	AAGACAGTGGCTGAGTGGCTCTCTATTAATCACTGACTAAGAAAGAGAAACTATTAAG	1437
OY	481	SerIeuValaRgArGSerTyTyRArgRArgGlyIySserGlnGlnGlnGlnGlnGln	500
Db	1438	AGCTTGATGAGCGGAGCTATGTGGCGCGCGGCAAGAGCAGACACAAACMACAGCAGAG	1497
OY	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnIuProMetProArgSerSerGlnGluIuIyS	520
Db	1498	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCGGAGCAGCAGGAGGAGAA	1557
OY	521	AspGluIySgIuIySgIuIySgIuAlaGluIySgIuGlnGlnIuIySProGluIuAlaGluAsn	540
Db	1558	GATGTGAAGAGAGAAAGAAAGAGGCGGAGAAAGAGAGAGAGAAACCGGAGGTGAGAAAC	1617

QY	541	AspIySGIuSpIeuleuLySGIuLyvThrAspAspThSergIyGluAaPaMaAspGIu	560
Db	1618	GACAAAGAAAGCCTCTCCAAAGAAAGACAGACACCTCCAGGGAGGACACAGCAG	1677
QY	561	LySGIuAaIaValaAserLySGIyArgLyvThraIaAsnSergIuIyArgLyvGIy	580
Db	1678	AAGAGGCTGTGGCTCCAAAGGCGCGAAACTCCAAACAGCCAGGGAGAAAGCAAGGC	1737
QY	581	ArgIleThraSermEtLaAsnGIuAaAsnSergIuGIuAaIleThrProGIuIn	600
Db	1738	CGATATCACCCGCTCAATGGCTTAATGAAGCCCAACGCAAGAGGCGCATACCCACAG	1797
QY	601	SerIaGIuLeuAaIaSerMetGIuLeuAsnGIuSerAtrGTPrThrGIuGIuMet	620
Db	1798	AGCCCGCAAGCTGGCTCCATGAGACTGAATGAGATTCTGGCTGACAGAAAGAAATG	1857
QY	621	GIuThraIaLyvLySGIyLeuLeuGIuInHISLyvArgAnThrPserAaIleAaIyMet	640
Db	1858	GAAACAGCCAAAGAAAGGCTCTCTCGAAACAGCGCGCAACTGTGTGGCATATCGCCGGATG	1917
QY	641	ValGIserLyvThraIaSerGIuCyLevAsnMetYrPheAnTyLyLyvArgGIuIn	660
Db	1918	GTGGGCTTCCAGACTGTGTGGCTAGTGAAGACTTCTTCACTTCACTACAGAGAGCAG	1977
QY	661	AsnLeuAspGIuIleLeuGIuInHISLyvLeuLyvMetGIuLyvGIuArgMaAaArg	680
Db	1978	AACCTCGATGAGATCTTGACAGCACACAAGCTGAAGATGAGAAAGAGAGAAAGCGCCGG	2037
QY	681	ArgLyvLyvLyvLyvAaIaProAaIaAaIaSerGIuGIuAaIaPheProProvaIyAaI	700
Db	2038	AGGAAGAAAGAAAGAGCGCGCGCGCGCGCACAGAGAGGGCTGATTCGCGCCGTGTG	2097
QY	701	GIuAspGIuGIuMetGIuAaIaSerGIyValaSerGIyAsnGIuGIuMetValGIuGIu	720
Db	2098	GAGATAGAGAGATGAGAGCGCTCGGGCGTGAACGGAAATGACGAGATGTGTGAAGAG	2157
QY	721	AlaGIuAaLeuHISaIaSerGIyAsnGIuValProArgGIyGIyCysSergIyProAaIa	740
Db	2158	GCTGAA-----GCC	2166
QY	741	ThyValaAsnAsnSerAAspThGIuSerIleProSerpProHISThrGIuAaIaAaLyS	760
Db	2167	ACTGTCAACACAGCTCAGACACCGAGAGACTCCCTCTCTCTCACTGAGCGGCCAAG	2226
QY	761	AspThrGIyGIuAsnGIyProLyvProProAaIaThrLeuGIyAaAspGIyProProPro	780
Db	2227	GACACAGGGCGAATGGGCGCCAAACCCCGACCGCTGGCGCGCGACGGGCGACCCCA	2286
QY	781	GIyProProThrProProPro-ArgaGmThrSerArgAaIaProIleGIuProThrProAaIaSe	800
Db	2287	GGGCGACCCACCCACCGACCGGAAGACATCCGGCC-CCACTAGATCCACCCGGGCTC	2345
QY	801	GIuAaIaThrGIyAaIaProThrProProProAaIaProProSerProSerAaIaProProPr	820
Db	2346	TGAAGCCACCTTAAGCGCCCTTAGCCGCCACCGACCGACCCCAATTTCTCTTCACTCTCC	2405
QY	820	OvaIyAaIaProLyvGIuLyvGIuGIuGIuThraIaAaIaAaProProvaIaGIuGIuGI	840
Db	2406	TGTGTCTCCCAAGAGAGGAGAAAGAGAGAGAACCGACAGCGGCCCCAGGTGAGAGAGG	2465
QY	840	YGIuGIuGIuInLyvProProAaIaAaIaGIuGIuLeuAaIaValaAspThrGIyLyvAaIaGIuGI	860
Db	2466	GAGAGAGCAAGCCCCCGCGGCTGAGAGACTGGCATGTGACACAGAGAAAGCGCGAGGA	2525
QY	860	uProvaIaLyvSergIyCysThrGIuGIuAaIaGIuGIuGIyProAaIaLyvGIyAspAaI	880
Db	2526	GCCGCTCAAGAGCGAGTGCACAGAGAGAAAGCGAGAGAGGCGCGCGCAAGGCGCAAGCGC	2585
QY	880	AGIuAaIaAaIaGIuAaIaThraIaGIuGIyAaIaLeuLyvAaIaGIuLyvLyvGIuGIyGIaSe	900
Db	2586	GGAGGCGCCCTAGGCGCAAGCGCGAGAGGCGCTCAAGCAAGAAAGAGAGGCGCGGAG	2645
QY	900	IGIyArgAaIaThraIaLyvSerSergIyAaIaProGIuAspSerAAspSerAaIaTh	920

Db 2646 CGGAGGCGCCACAGCCAGAGACTCGGGGCGCCCCAGAGCAGCAGCTCAGTGGCCAC 2705
Qy 920 rCysSerAlaAspGluValaAspGluValaGluGluValaAspLysValaMetLysSerPr 940
Db 2706 CTGCAATGACAGACAGGATGATGAGGCGGAGGCGGCGACAAAGAACCGGCTGCTGCC 2765
Qy 940 oArpProSerLeuLeuThrProThrGluYAspProArgAlaAsnAlaSerProGlnLysPr 960
Db 2766 AAGGCCAGGCTCTCAACCCCGACTGGAGACCCCGGGCCAAATGCTTACCCCAAGAGCC 2825
Qy 960 oLeuAspLeuLysGlnLeuLysGlnAArgAlaAlaAlaLeuProProIleGlnValThrLy 980
Db 2826 ACTGAGCCTGAAAGAGCTGAAAGCAGCGGCTGCCATCCCGCCCATCCAGCTCACCAA 2885
Qy 980 sValHIEGluProProArgGluAspAlaAlaProThrLysProAlaProAlaProPr 1000
Db 2886 AGTCAATGAGCCCCCGGAGAGAGCAGCTCCACCAAGCAGACTCCCGCAGCCCAAC 2945
Qy 1000 oProProGlnAsnLeuGlnProGlnSerAspAlaProGlnGlnProGlnLysSerProAr 1020
Db 2946 GCGACCGCAAAACCTGACGCGGAGAGAGAGCCCTCAGAGGCTGGCAGCAGACCCCG 3005
Qy 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3006 GGGCAAGAGAGAGAGCGCGGACCCCGCGCAAGAGAGCAGAGAAAGCTGTGTTCTT 3065
Qy 1034 -----AlaPheAlaAlaGluValaGluLysLeuProGluYAspProProCysTrpThrSerG 1052
Db 3066 CCGAGCTTGGACGCGCCAGAGCGCCAGAGCTGCGGAGACCCCTTGCTGAGACTTCGG 3125
Qy 1052 YLeuProPheProValaProProArgGluValaIleLysAlaSerProHisAlaProAspPr 1072
Db 3126 CTGGCTTCCCGCTGCGCCCCCTGAGGTGATGAAGGCTCCCGCATGCGCCGGAGACC 3185
Qy 1072 oSerAlaPheSerArgAlaProProGluYHisProLeuProLeuGluLeuHisAspThrAl 1092
Db 3186 CTGAGCTTCTCTCAACGCTCAACCTGTCACCCAGCTCCCTGGGCTCCCATGACATGC 3245
Qy 1092 sArgProValaLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3246 CCGGCGCGTCTGCGCGGCCACCCACCATCTCCAAACCGCTCCCTCATCTCTGTC 3305
Qy 1112 aLysHisProSerValaLeuGluValaArgGlnIleGluValaIleSerGlnIleMetSerValG 1132
Db 3306 CAAGGACCCAGGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGTCGCTCA 3365
Qy 1132 pLeuHisValaProTyrSerGluHisAlaLysAlaProValaGluProValaThrMetGlyE 1152
Db 3366 GCTTCAAGCTCCGTACTCAGAGCATGCCAAAGCCCGGTGGGCTGTCTACATGGGGCT 3425
Qy 1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGluValaLysGlnGluGlnE 1172
Db 3426 GCGCTGCGCATGAGACCCCAAAAGCTGGACCCCTTACGGAGTGAAGAGGAGCGCT 3485
Qy 1172 uSerProArgGluGlnAlaGluProProGlnSerLeuGluValaProThrAlaGlnGlu 1192
Db 3486 GTTCCCAAGGGGCGAGGCTGGGCGCACCGAGAGCTGGGGGTGCCACAGCCCAAGAGGC 3545
Qy 1192 aSerValaLeuArgGluYHisAlaLeuGluYSerValaProGluYSerIleThrLysGluY 1212
Db 3546 GTCCGCTGCTGAGAGGAGCAGCTCTGGGCTCAGTTCCGGGCGAAGCATCAACAAAGCAT 3605
Qy 1212 eProSerThrArgValaProSerAspSerAlaIleThrThrArgGluYSerIleThrHisG 1232
Db 3606 TCCCAAGACACGGGATGCTCTGAGACAGCGCATACATACCGGGCTCCATCAACCGG 3665
Qy 1232 YThrProAlaAspValaLeuTyrLysGluYHisIleThrArgIleIleGluYHisAspSerPr 1252
Db 3666 CAGGCAAGCTGAGCTCTGTACAAAGGCAACATCAACAGATCATCGCGCAGAGCAGCCC 3725
Qy 1252 oSerArgLeuAspArgGluYArgGluAspSerLeuProLysGluYHisValaIleTyrGluG 1272

Db 3726 GAGTCGCTTGGACCGCGCGGAGAGCAGGCTGCCCAAGGCGCACATCAAGAAAG 3785
Qy 1272 YLysLysGluYHisValaLeuSerTyrGluGluYHisMetSerValaThrGlnCysSerLysG 1292
Db 3786 CAAGAGGGCCAGCTCTTGTCTTATGAGGTGGCATGTCTGTACCCAGTCTCCAAAGGA 3845
Qy 1292 uAspGluYArgSerSerSerGluProProHisGluThrAlaAlaProLysArgThrTyrAs 1312
Db 3846 GAGCGCAGAAAGAGCTCAGAGACCCCGCCATGAGAGCGCGCCCGCCCAAGGCACTTATGA 3905
Qy 1312 pMetMetGluGluYArgValaGluYArgAlaIleSerSerAlaSerIleGluGluYMetG 1332
Db 3906 CATGATGAGAGGCGCGCTGGGACAGAGCATCTCTCAGCCAGCATCAAGGCTCTCATGG 3965
Qy 1332 YArgAlaIleProProGluThrHisSerProHisHisLeuLysGluGluHisHisIleAr 1352
Db 3966 CCGTGCCATCCCGCGAGGACACAGCCCGCCACACTCAAAAGAGAGCAGCATCCG 4025
Qy 1352 gGlySerIleThrGlnGluYLeuProArgSerTyrValaGluAlaGlnGluAspTyrLeuAr 1372
Db 4026 CCGGTCCATCAACAAGGATCCCTGTGCTCTAGTGAAGGCAAGAGGACTTACCTGGC 4085
Qy 1372 gArgGluAlaLysLeuLysArgGluGluYThrProProProProProSerArgAs 1392
Db 4086 TCGGAGGCGCAAGCTCTTAAAGCGGAGGAGCAGCGCTCCCGCCCAACCGCTCAACGGGA 4145
Qy 1392 pLeuThrGluAlaTyrThrThrGlnAlaLeuGluYProLeuLysLeuYAspProAlaHisG 1412
Db 4146 CTGACCGAGCGCTTACAGAGCAGAGCGCTGGGCGCCCTTAAGCTGAAGCGCGCCATGA 4205
Qy 1412 uGluLeuValaAlaThrValaLysGluValaGluYArgSerIleHisGluIleProArgGluG 1432
Db 4206 GGGCTGTGTGCCACGGTGAAGAGAGCGGGCCCTCATCATGAGATCCCGCGAGGA 4265
Qy 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluYSerIleTh 1452
Db 4266 GCTGGCGCACACCGCGAGCTGCCCTGGCGCGCGCGCTCAAGAGAGGCTCATCAC 4325
Qy 1452 rGlnGluYThrProLeuLysThrThrArgValaAspThrThrGluYSerLysLysHisAs 1472
Db 4326 GAGAGGACCCCGCTCAAGTACAGACAGCGCGGTCCACATGCTCCCAAAAGACAGA 4385
Qy 1472 pValaArgSerLeuIleGluYSerProGluYArgThrPheProProValaHisProLeuAspVa 1492
Db 4386 CGTACGCTTCCATATGGGAGCGCCCGCGGAGGATTCACACCGGTGACCGGTGATGT 4445
Qy 1492 pMetAlaAspAlaArgAlaLeuGluValaGluYArgGluYSerLysSerArgPr 1512
Db 4446 GATGGCGGACGCGCGGGCACTGGAAAGTGGCTCTCAAGAGAGGCTTGAAAGCGCGGCC 4505
Qy 1512 oGluYThrAlaSerSerSerGluYSerIleHisAlaArgGluValaProValaIleValaProG 1532
Db 4506 AGGAGACCGCAGAGCTCGGGGGGCTCAATGGCGCGGCGCGCGGCTATTTGCTTGA 4565
Qy 1532 uLeuGluYLysProArgGlnSerProLeuThrTyrGluAspHisGluValaProPheAlaG 1552
Db 4566 GCTGGGTAAAGCCCGGCAAGAGCCCGCTGACATGAGACAGAGGGGCAACCTTTCGG 4625
Qy 1552 YHisLeuProArgGluYSerProValaThrMetArgGluProThrProArgLeuGlnGlu 1572
Db 4626 CCAACCTTCCACGAGGTTCGCGCGTGAACATGCGGAGAGCCACCGCGCTTGAGAGGG 4685
Qy 1572 YSerLeuSerSerSerSerValaSerGlnAspArgLysLeuThrSerThrProArgGlu 1592
Db 4686 CAGCTTGTGTCAGAGAGGATCCAGAGCAGAAAGCTGACGTGACGCTGTGATGAT 4745
Qy 1592 eAlaLysSerProHisSerThrValaProGluHisHisProHisProIleSerProTyrG 1612
Db 4746 CCGCAAGTCCCGGCAAGACAGCGGTGCCAGAGACACCAACCCCATCTCGGCTATGA 4805
Qy 1612 uHisLeuLeuArgGluValaSerGluValaLysLeuTyrArgSerHisIleProLeuAlaAsp 1632
Db 4806 GCACCTGTTCCGGGCGTGAAGTGGCGTGAACCTGTATGCAAGCAGATCCCGCTGGCTT 4865

[illegible]

QY	1992	aThrlIalalarghrhProalalybaamleuA1aProhiShiIalaserProaapPro	2012
Db	5946	CACCAATGCCCGCACCCCTCGGAAGAACTTGCACTTCACACGCCGACCCGGACCCGCC	6005
QY	2012	oAlaProProalaserAlaserAapProhiSaRGJuluYThriInseryProphSe	2032
Db	6006	GGCGCACTCGCTCGGCTCGACCCGGACCGGGAAGAAAGACTCAAGATAAACCTTTTC	6065
QY	2032	rIleGIuIuleuIuleuAargserleuGIYThriGIYserGerYrsrProGIuGI	2052
Db	6066	CATCCAGAACTGGAACTCGTTCTCGGTATTACACGGCAGCACTACACCCCGAAGG	6125
QY	2052	YAlaGIuProalSerProValSerSerProSerleuThriSaBpYsgIYleuProly	2072
Db	6126	GGTGAAGCCGCTCAGCCCTGTGACTACCAAGCTGACCCACGACAAGGGGCTCCCAA	6185
QY	2072	ShIleuGIuGIuleuAaplySerHShleuGIuGIYIuleuAgrpPolyGIuProGI	2092
Db	6186	GCACCTGGAAAGCTCGACAAGAACCACTTGAGAGGGAGCTGGGCCCAAGCAAGCCAGG	6245
QY	2092	YProVallysaGIuGIYGIuAlaAhiSaIeueProhiSleuAgrpProleuProGIuSe	2112
Db	6246	CCCCGTGAAGCTTGGCGGGAGGCCGCCCACTCCACACTGGCGCGCTGCTTGAGAG	6305
QY	2112	GIuProSerSerSerProleuGIuGIuThraIaProGIYAllySGIYhiSGIuAryA	2132
Db	6306	CCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGGCCAGGGGTCAAAAGTCCACAGCGGT	6365
QY	2132	IYalThrlleuAagIuIiSleSGIuIalIeThGIuAaplyrThraGhiShIaPr	2152
Db	6366	GGTCACTCGGCCACGACATCAGTAGAGTATCACAGGACTACACCCGGCACACCC	6425
QY	2152	oGIuGIuInseryAlaProleuProAlaProleuTySerPheProGIYAlaSerCyPr	2172
Db	6426	ACAGAGCTCAGGGGACCCCTGGCCGCCCTCTACTCTTCCGTGGAGGCACTGGCC	6485
QY	2172	oValleuAapleuAarghrProProSerAapleuTyIleuProProPolaaphIsgIYAl	2192
Db	6486	CGTCTGACCTCCGCGCGCCACCCAGAGACTTACTCCCGGCCCGGACCAATGGTGC	6545
QY	2192	aProIalargIYserProhiSaserGIuGIYGIYsaGrSerProGIuProAmluYTh	2212
Db	6546	CCGGCCCGTGGCTCCCCCACAAGCGAAGGGGGCAAGGTCTCCAGAGCCMAACAGAC	6605
QY	2212	rSerValleuGIYGIYGIuAapGIYIleGIuProValSerProProGIuGIYmetTh	2232
Db	6606	GTCGGCTTGGGTGGTGGTAGAGACCGGATTAAGAACTGTGTCCCAACCGGAGGGCATAC	6665
QY	2232	GIuProGIYhiSserAargserAlaValTyProleuIuIuYrAargBpGIYGIuGIuTh	2252
Db	6666	GGAGCGAGGGACCTCCCGAGTGTGTGTATCCCTGTGTACCGGGATGGAGAACAC	6725
QY	2252	rGIuProSerAargmetGIYserTySerProGIYAsnThSerGIuProProAlaPhePh	2272
Db	6726	GGAGCCCAAGAGTGGCTCCAGTCTCCAGGACAACAAGCCAGCGCCAGCCCTTCTT	6785
QY	2272	eSerIYleuThrgIuSerAaSerAlaMetVallySerIuYsgIuIleuIuIeInsly	2292
Db	6786	CAGCAAGCTGACCGAGAGCAACTCCGCATGTGTCAAGTCCAAAGAACGAAGATCAACA	6845
QY	2292	ShIYleuAasnthriSaBaAryaangIuProGIuTyAsnIleSerGIuProGIYThGI	2312
Db	6846	GAAAGCTGAACACCAACAACCGGAATGAGCTGAATCAATTCAGCCAGCTGGGACGA	6905
QY	2312	uIlePheAaMetProAlaIleThrGIYThGIYleuMetThrTyAargSerGIuAlaVa	2332
Db	6906	GATCTTCAATATGCCCGCATCACCGGAACAGGCTTATGACTATAGAAACCCAGGCGGT	6965
QY	2332	IGIuGIuIiSaIaserThraMetGIYleuGIuAlaIleIleGrlySaIaleuMetGI	2352
Db	6966	GCAGGAATATCCACGACCAACATGGGGCTGAGAGCCATAATGTGAAGAGGCACTATAGG	7025
QY	2352	YIleYrAapGIuProGIuGIuSerProProleuSerAlaSnAlaPheAaProleuAs	2372

Db 7026 TAAATATACCAAGTGGAAAGTCCCGCCGCTCAGCCCAATGCTTTAAACCTCTGAA 7085
Qy 2312 nAlSerAlaSerLeuProAlaAlaMetProIleThrIaAlaAspGlyAAspSerAAspHi 2392
Db 7086 TGGCAGTCCAGCCCTGGCCGCTGCTATCCCTAAACCGCTGTCAGAGAGAGAGCA 7145
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyAlaValSerGlyAAspProSerSerA 2412
Db 7146 CACACTACCTCGCCAGGTGGCGGGAGAGGCCAAGTCTCTGGCAACCCAGAGCCG 7205
Qy 2412 GLyAlaIAspSerProAlaProGlyLeuAlaSerGlyAAspArgProProSerValSerSe 2432
Db 7206 AAAAGCCAAAGTCCCGCGCGGCTGGCATCTGGGAGACCGGCAACCTCTGTCCTC 7265
Qy 2432 rValHisSerGlyGlyAAspCyAspAAspArgThrProLeuThrAAspValTTPGluAs 2452
Db 7266 AGTCACTCGAGAGAGAGACTCAACCGCGGACCGCTCAACACCGCTGTGGAGAG 7335
Qy 2452 PaRGProSerSerAlaGlySerThrProPheProTyraProLeuIleMetArGLeuG 2472
Db 7326 CAGGCCCTCGTCCGAGGTTCCAGCCATTCCTCTACAAACCCCTGATCATGGGCTGCA 7385
Qy 2472 nAlGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7386 GCGCGGTGTGATGAGCTTCCACCCGCGGCTCCCGCGGAGCGGCGGCGGCTCGC 7445
Qy 2492 aGlyProHisHisAlaThrAspArgIuGluProIleProLeuLeuCySerGlnTyrGluTh 2512
Db 7446 TGGCCCCCAGCCAGCTGGGAGAGAGGCCAAGCCACTGCTGTGTCAGATAGAGAC 7505
Qy 2512 rLeuSerAAspSerGlu 2517
Db 7506 ACTCTCGAGAGCCAG 7521
RESULT 10
ID ACA62249 standard; cDNA; 8686 BP.
XX ACA62249;
XX ACACA62249;
XX 12-AUG-2003 (first entry)
XX cDNA encoding human nuclear receptor corepressor SMRte.
XX Human; sex; gene; SMRte; nuclear receptor corepressor; gene therapy;
XX tissue typing; cancer.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 157..7680
XX FT /*tag= a
XX FT /product= "SMRte"
XX US2003027137-A1.
XX PD 06-FEB-2003.
XX PF 27-MAR-2001; 2001US-00819104.
XX PR 29-MAR-2000; 2000US-0193138P.
XX PA (CHEN/) CHEN J D.
XX Chen JD;
XX WPI; 2003-466139/44.
XX DR P-PSDB; ABU61812.
XX PT New SMRte proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.

XX Claim 2; Page 32-41; 90pp; English.
PS The invention relates to an isolated SMRte nucleic acid molecule. The
XX nucleic acids are useful in gene therapy, as hybridization probes for
CC identifying SMRte-encoding nucleic acid molecules and as primers for
CC amplifying of SMRte nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRte are useful for treating or
CC preventing a condition associated with aberrant SMRte protein or nucleic
CC acid expression or activity, such as cancer. The present sequence
CC represents cDNA encoding the human nuclear receptor corepressor SMRte
XX Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 8 Gaps: 3
US-09-522-753-5 (1-2517) x ACA62249 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 157 ATGTGGGGCTCCACAGCCGTGGGCACAGACGTAAGGGGCACTAGACCCCGTACCGG 216
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 217 CCCACAGCCTTTCCTACCAAGTGCAGATGCCCGGAGCGACAGAGCTGGGCTCCTG 276
Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 277 GAGTACAGACACCACTCCCGGACTATGCTCCCACTGTGCGCGGCTCATATATCCAG 336
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAAspGlnArgSerGln 80
Db 337 CCCAGCGGGAGAGGCCCTCCTGCTGTGTGATTCAGCCCGGGAATGAACGGTCCAG 396
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleSerGluMet 100
Db 397 GAGCTCCACTGGCGGCCAGAGTCCCACTACTGCTGCCAGGCTGGGGAAGTCAGAGATG 456
Qy 101 GluPheIleGluSerIleAspArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 457 GAGTTCATTAAAGCAAGCCGCTCGGTAGAGTGTGCTGACCCCTGCTGGCACCG 516
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140
Db 517 TCACCCCTGCTGGGCACGGGCCAGCCGTGGGAGTCTGAACACCTCAAGAGACCGTAGC 576
Qy 141 LeuThrGlyIleLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGGAGCCCGGCACTGACCTGTAGCTG 636
Qy 161 GluLeuValProProArgLeuSerIleGluGluLeuIleGlnAspMetAspArgValAsp 180
Db 637 GAGCTGTGCGCCAGCCAGCTGTCCAAAGAGAGACTGATCCAAACATGACCCGCTGAGC 696
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerIleLeuIleValIleGlnGlnLeu 200
Db 697 CGAGAGATCACCATGTAAGAGAGAGATCTTAAGCTGAAGAAAGAGCAACAGCTG 756
Qy 201 GluGluGluAlaAlaIleAspProProGluProGluIleProValSerProProIleGlu 220
Db 757 GAGAGAGAGGCTGCCAAGCGCGCCGAGCTGAGAAACCCGCTGTACCGCGCCATTCGAG 816

QY	221	SetrYvHsAaGserLeuValGlnIleLeIytrAepGlsuAmaAqVlybAAGluA	240
Db	817	TCGAAGAACCCGACGCTGTGTCAAGTCACTTACGACGAGAACCCGGAAGAAGCGTGAACCT	876
QY	241	AlAhIsArqIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyraSngInPro	260
Db	877	GCAATCGGATTTCTGGAAAGGCGCTGGGGCCCAAGGTGGAGCTGTCCGCTTACACCAAGGCC	936
QY	261	SezApbThraArgInTyHieGluAsnIleLysIleasnGlnAlaMetArgLysIleu	280
Db	937	TCCCAACACCCCGCAGTATCATGAGAACATCAAAATTAACCAAGCGCATCGGAAGAAGCTA	996
QY	281	IleLeuTyRPhLyArqArqAsnHsAlaArgLysGlnTyRPhLyGlnTyRPhLyGln	300
Db	997	ATCTTGATCTTCAAGAGGAGGAATACCTCGGAAACATGGAGCAACAAGTTCGTCAG	1056
QY	301	ArgTyraepGlnLeuMetGluAlaLeuGlnLysLysValGluArgIleGluAsnAsnPro	320
Db	1057	CGCTATGACCAAGCTCATATGAGGCGCTCGGAGAAAGAGTGGAGCGCATGAGAACCAACCC	1116
QY	321	ArgArqArqAlaLysGluSerLysValArgGlnTyTyRglnLysGlnPheProGlnIle	340
Db	1117	CGGCGGCGGGCCAAAGAGAGCAAGGTTGCGGAGTACTACGAAAGCAAGTCCCTGATGTC	1176
QY	341	ArgLysGlnArqGlnLeuGlnGlnLuarMetGlnSerArqValGlyGlnArgGlySerGly	360
Db	1177	CGCAAGCAAGCCGAGCTGACAGAGCGCATGCG--AAGGTGGGCGACGGGGCGTGGG	1233
QY	361	LeuSerMetSerAlaAlaArgSerGlnHsGlnValSerGlnIleIleAspGlyLeuSer	380
Db	1234	CTGTCTCATGTGCGCCCGCCGACGAGACGACGAGGTGTACAGATCATGATGGCTCTCA	1293
QY	381	GluGlnGlnAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyR	400
Db	1294	GAGCAGAGGAAGCTGAGAGAACAGATGGCCAGCTGGCCGATGATCCGCCATGCTGTAC	1353
QY	401	AspAlaAspGlnGlnArqIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet	420
Db	1354	GACCTGTACCAGCGCGCATCAAGTTCATCAACATGAACGGGGCTTATGGCGACCCCATG	1413
QY	421	LysValTyRlysAspArqGlnValMetAsnMetTrpSerGlnGlnGlnLysGlnTyRPh	440
Db	1414	AAAGTGTACAAAGACCGCCAGGCTCATGAACTGTGAGTGAACAGAGAAAGAGACCTTC	1473
QY	441	ArgGlnLysPheMetGlnHsProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg	460
Db	1474	CGGAGGAAGTTCAATGACACATCCCAAGAACTTTGGCTGATGCAATCTCTCGAGAGG	1533
QY	461	LysThrValAlaGluCysValLeuTyTyTyRLeuThrLysLysAsnGluAsnTyRlys	480
Db	1534	AAAGCAGTGGCTGAGTGGCTCTCTATTACTTACCTGAAGAAAGAAATCAATATAG	1593
QY	481	SerLeuValArgArqSerTyRArqArqArgGlyLysSerGlnGlnGlnGlnGlnGln	500
Db	1594	AGCCTGTGAACGGAGCTATGTGGGCGCGGAGCAAGACCAAGCAACAACGACGAG	1653
QY	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnLys	520
Db	1654	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1713
QY	521	AspGlnLysGlnLysGlnLysGlnAlaGlnLysGlnGlnGlnLysProGlnValGluAsn	540
Db	1714	GATGAGAAGAGGAAGAAAGAGCGGAGAAAGAGAGGAGGAGGAGGAGGAGGAGGAG	1773
QY	541	AspLysGlnAspLeuLeuLysGlnLysThrPheAspThrSerGlyLysAsnAspGln	560
Db	1774	GACAAAGGAAGCTCTCTCAAGAGAGCAACACGACCTTAGGGGAGAGCAACGACAG	1833
QY	561	LysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnLysArgLysGly	580
Db	1834	AAAGAGGCTGTGGCTCCAAAGCGCCGCAAACTGCCAACAGCCAGGAGAGACGCAAAAGC	1893
QY	581	ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlnGlnAlaIleThrProGlnGln	600

[illegible]

Db 2922 AAGCCAGCCTCTCAACCCGAGTGGACCCCGGGCCAATGCTTCAACCCAGAAACC 2981
 Qy 960 OLeuApeLeuYsglInleuYsglInArgAlaAlaIleProProIleGluValIleThrLy 980
 Db 2982 ACTGAGCTGAAGAGCTGAAGAGCAGCGGCTGCTCCATCCCCCATCCAGGTCAACCA 3041
 Qy 980 aValHleGluProProArgIuAspAlaAlaProThrLyAspProAlaProProAlaProPr 1000
 Db 3042 AGTCATAGAGCCCCCGGAGAGCAGCGCTCCCAACCAAGCAAGTCCCGACGCCCAACC 3101
 Qy 1000 oProProGluAsnLeuGlnProGluSerAspAlaProGluGlnProGluYserSerProAr 1020
 Db 3102 GCGACCCGAAAACTGACGCGGAGAGGAGCGCCCTCAGCAGCCTGGCAGCACGCCCGG 3161
 Qy 1020 gGlyYserSerArgSerProAlaProProAlaAspLyseGlu----- 1033
 Db 3162 GGGGAAGAGCAGAGCCCGGACCCCGCGGACAGAGGAGGACAGAAAGCTGTGTTCTT 3221
 Qy 1034 ----AlaPheAlaAlaGluAlaGluYleuProGluYAspProProCystrProThrSerG 1052
 Db 3222 CCCAGCCTTGGCAGCCGAGGCCAGAAAGCTGCTGGGGACCCCTTGCTGACTTCGG 3281
 Qy 1052 YleuProProProAlaProProAlaArgIuValIleYleuAlaSerProHleAlaProAspPr 1072
 Db 3282 CTTGCCCTTCCCGCTGCCCGCCCGTGAAGTGATCAAGGCTCCCGCATGCTCCGAGACC 3341
 Qy 1072 oSerAlaPheSerTyraAlaProProGluYHleProleuProleuGlyleuHleAspThrAl 1092
 Db 3342 CTCAGCCTTCTCTCAAGCTCACTGATCAGCCACCTGCTCCGAGCTCCATGACACTGC 3401
 Qy 1092 aArgProValleuProArgProProThrIleSerAsnProProProleuIleSerSerAl 1112
 Db 3402 CCGGCGCTGCTGCGCGGCCACCAACATCTCAACCCGCTCCCTCATCTCTCTGCG 3461
 Qy 1112 alyHleAspSerSerValleuGluValIleGlyAlaIleSerGlnGlyMetSerValG 1132
 Db 3462 CAAGCACCCAGAGCTCTCGAGAGCAAAATAGTGCCATCTCCCAAGAAATGTGCGTCCA 3521
 Qy 1132 nLeuHleValProTySerGlnHleAlaYleAlaProValGlyProValIleThrMetGlyLe 1152
 Db 3522 GCTTCAAGTCCCTGACTCAAGCATGCGCAAGGCCCGGAGGCTGTCACATGAGGCT 3581
 Qy 1152 uProleuProMetAspProlyYleuAlaProPheSerGlyValYleGlnGlnGlnLe 1172
 Db 3582 GCCCCTGCCATGAGCCCAAAAGCTGGACCTTCAGCGAGTGAAGAGAGGAGCGCT 3641
 Qy 1172 uSerProArgGlyGlnAlaGlyProProGluYserLeuGlyValProThrAlaGlnGluAl 1192
 Db 3642 GTCCCCACGGGGCCAGGCTGGGCGACCGAGAGGCTGGGGGTGCCACAGCCGAGAGGC 3701
 Qy 1192 aSerValleuArgGlyThrAlaLeuGlyserValProGlyGlySerIleThrLyseGlyI 1212
 Db 3702 GTCCGAGTCAAGAGGAGCACTGAGGCTCAAGTCCGAGGAGAGCATCAACCAAGGAT 3761
 Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrHleG 1232
 Db 3762 TCCAGACACGCGGTGCTCGACAGCGCATCACTACCGGGCTCCATCAACCCAGG 3821
 Qy 1232 YThrProAlaAspValleuTyrylyGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
 Db 3822 CAGCGCCAGCTGAGCTCTGTACAAAGGAGCACCATCAACAGATCATCCGCGAGGACGCC 3881
 Qy 1252 oSerArgLeuAspArgGlyValArgIuAspSerLeuProYleuGlyHleAlaIleTyryGlu 1272
 Db 3882 GAGTCCGTTGACCGCGCGGAGAGACAGCTGCCAAGGGCCAGCTCATCGAAGG 3941
 Qy 1272 YlyYleGlyHleAlaLeuSerTyryGluGlyMetSerValIleThrGlnCyserLyseG 1292
 Db 3942 CAAGAAAGGCCACGCTTTGTCTATGAGGAGGAGCATCTGTGACCAAGGCTCCAAAGGA 4001
 Qy 1292 uAspGlyYArgSerSerSerGlyProProHleGluThrAlaAlaProLyAspArgThrTyra 1312
 Db 4002 GGAAGGAGAGGAGCTCAAGGACCCCTCCATGAGAGCGCGCCCGCAAGGCGACCTATGA 4061

Qy 1312 pMetMetGlyGlyYArgValGlyYArgAlaIleSerSerAlaSerIleGlyLeuMetG 1332
 Db 4062 CATATGAGAGGCGCGCTGGGAGAGCCATCTCTCAAGCAGCATCAAGGCTCATAGG 4121
 Qy 1332 YArgAlaIleProProGluYArgHleSerProHleHleLeuYsglGlnHleHleAla 1352
 Db 4122 CCGTGCATCCCGCGGAGGACACAGCCCAACCACTCAAAAGAGAGCACCATCCG 4181
 Qy 1352 gGlySerIleThrGlnGlnIleProArgSerTyraGluAlaGlnGlnHapTyryLeuAr 1372
 Db 4182 CGGCTCATACACAAAGAGATCCCTGCTCTTACGTGAGGACAGAGAGCATCTGCGG 4241
 Qy 1372 gArgGluAlaYleuLeuYsaArgIuGlyThrProProProProProProSerArgAs 1392
 Db 4242 TCGGAGAGCCAGCTCTTAAGCGGAGGAGCAGCTCCCGCCCAACCGCTCAACGGGA 4301
 Qy 4302 CTGACCGAGGCTTAACAAGAGCAGGCTGGGCCCCCTGAAGCTGAAGCCGCGCATGA 4361
 Db 4362 uGlyLeuValAlaThrValYsglAlaGlyYArgSerIleHleGluIleProArgGlu 1412
 Qy 1412 uGlyLeuValAlaThrValYsglAlaGlyYArgSerIleHleGluIleProArgGlu 1432
 Db 4362 GAGCTGTGTGCCACGGTGAAGAGGAGGCGCTCCATCATGAGATCCCGCGAGGA 4421
 Qy 1432 uLeuArgHleThrProGluYleuProleuAlaProArgProleuYsglYserIleTh 1452
 Db 4422 GCTGCGGACACCGCCGAGCTGCTGCTGCGCCCGCGGCTCAAGAGAGGCTCATCAC 4481
 Qy 1452 rGlnGlyThrProleuYsTyraAspThrGlyAlaSerThrGlyYserLyseYleHleAs 1472
 Db 4482 GCAAGGACCCCGCTCAAGTACAGACCGCGGCTCCACCATGAGTCCAAAAAGACAGA 4541
 Qy 1472 pValArgSerLeuIleGlySerProGlyYArgThrPheProProValHleProleuAspVa 1492
 Db 4542 CGTACGCTCCCTATCGGAGCGCCCGGCGAGCGTCCCAACCGTGACCCGCTGAGTG 4601
 Qy 1492 lMetAlaAspAlaArgAlaLeuGluYArgAlaCyTyryGluGlySerLeuYsSerArgPr 1512
 Db 4602 GATGCGCGAGCGCCGGGCACTGAGACGTGCTCTCAAGAGAGAGCTGAAGAGCGGCC 4661
 Qy 1512 oGlyThrAlaSerSerSerGlyYserIleAlaArgGlyYalaProValIleValProG 1532
 Db 4662 AGGAGCCGCGAGAGCTCGGGGGCTTCATTTGCGGGCGCCCGCTATTGGCTGA 4721
 Qy 1532 uLeuGlyYleProArgGlnSerProleuThrTyryGluAspHleGlyYalaProHleAla 1552
 Db 4722 GCTGGGTAAAGCCGCGGAGAGCCCTGACCTATGAGGACACAGGGGACACCTTTGCCG 4781
 Qy 1552 YHleAspProArgGlySerProValIleThrMetArgGluProThrProArgLeuGlnGlu 1572
 Db 4782 CCACTTCCCAAGAGTTCGCCCTGTACATGCGGAGCCCAACCGCGCTGACAGAGGG 4841
 Qy 1572 YserLeuSerSerSerLyseYleuAlaSerGlnAspArgYleuThrSerThrProArgGlu 1592
 Db 4842 CAGCTTTGCTCAGCAAGGATCCCAAGAGCGAAAGCTGACGTGCGCTGTGATGAT 4901
 Qy 1592 eAlaYleSerProHleSerThrValProGluHleHleProHleProIleSerProTyryG 1612
 Db 4902 CGCAGAGTCCCGCAAGACACCTGCGCGAGCACCAACCAACCATCTCGGCTATGA 4961
 Qy 1612 uHleLeuLeuArgGlyValIleSerGlyValAspLeuTyraArgSerHleIleProleuAlaPh 1632
 Db 4962 GCACCTGCTTGGGGGTGAGTGGCTGTATCGAGGACATATCCCTTGCGCTT 5021
 Qy 1632 eAspProThrSerIleProArgGlyIleProleuAspAlaAlaAlaTyryLeuPr 1652
 Db 5022 CGAGCCCACTCATACCGCGGATCCCTTGAGAGGAGCGGCTGCTCATACGCGCC 5081
 Qy 1652 oArgHleAlaAlaProAspProThrTyryProHleYleuTyryProProTyryLeuIleArg 1672
 Db 5082 CGGACACTGTGCCCCCAACCCCACTTACCGGACCTGTACCAACCTTACTTCGCGG 5141

QY 1672 yTyPProAspThrAlaAlaLeuGluAsnArgInThrIleIleAsnAspTyrlleThrSe 1692
Db 5142 CTACCCCGACACGGCGCGCTGGAGAAACCGGACAGCATCATCATGACTATCACTACCTC 5201
QY 1692 rGInGImeThiHiAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgI 1712
Db 5202 GCGAGAGATGCACCAACAACGGCCACCGCATGGCCACGAGCTGATATGCTGAGGG 5261
QY 1712 yLeuSerProArgIuSerSerLeuAlaLeuAsnTyrlaAlaGlyProArgIyIleI 1732
Db 5262 CCTCTGCCCCGGAGATCCTCGCTGGACATCACTACGCTGGGGGTCCCGAGGACATCAT 5321
QY 1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
Db 5322 CGACCTGTCGCAAGTGCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5381
QY 1752 aThrAlaMetAspArgLeuAlaTyrlleuPProThraAlaProGlnProPheSerArgH 1772
Db 5382 CACCGCCATGACCGCTTGCCTACCTCCCGCCGCGCCCGCCAGCCCTTCAGCAGCCGCA 5441
QY 1772 sSerSerSerProLeuSerProGlyGlyProThriSleuThrLysProThrThrSe 1792
Db 5442 CAGCACTCCCACTCTCCCAAGAGGTCCACACACTTGAACAAACCAACCAACCACTC 5501
QY 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluArg 1812
Db 5502 CTCTGTCGAGCGGAGCGAGACCGGATCGAGACCGGAGACCGGAGATCGGAGCGGAGAAA 5561
QY 1812 sSerIleleuThrSerThrThrThrValGlnIleAlaProIleThrArgProGlyThrG 1832
Db 5562 GTCAATCTCACTCAACCAACGAGGAGAGACGACCATCTGGAGACTGTGTACAG 5621
QY 1832 uGlnSerSerGlySerSerSerGlySerSerGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCGAGACAGCGGAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5681
QY 1852 aSerThiSserThiAlaHiGlnHiSserProIleSserProArgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCACTCCCACTCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 5741
QY 1872 nGlnArgProSerValLeuHiAsnThrGlyMetLysGlyIleIleThrAlaValGluP 1892
Db 5742 GCGAGACCGAGTGTCTTCAACAACAGGATGAAGGATATCATCCGCTTGGAGCC 5801
QY 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIaTh 1912
Db 5802 CAGCAAGCCACAGGTCTCTGAGTGCACCTCACTCTCACTCTCACTCTCACTCTCACT 5861
QY 1912 rPheProProAlaThrHiSsCyAProLeuGlyGlyThrLeuAspGlyValTyrlProThrI 1932
Db 5862 ATCCCACTCCGACCACTGCGCACCTGAGGCGGACCTCGATGGGGTCTACCTTACCT 5921
QY 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGGAGCCGCTTGTCTGCTCAAGAGAGGCCCCCGGGTGCCTCGGCAAGGCGCCCG 5981
QY 1952 gAlaAspThrGlyHiAlaIleLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCAATGCTTCTCTCGCAAGCCCCCAGCCGCTCGGGGTGAGCCCGC 6041
QY 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHiAl 1992
Db 6042 CTCTCTCCCGCAGAGAGGCTCGAGAGCCCGGCCCCCTAGTGTCTCTGCTGCGCACGC 6101
QY 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHiSsAlaSerProAspProPr 2012
Db 6102 CACCATGCGCGCACCCCTGCGAAGAACTTGCACCTCAGCAGCAGCAGCAGCAGCAGC 6161
QY 2012 oAlaProAlaSerAlaSerAspProHiSsArgGluLysThrGlnSerLysArgPheSe 2032
Db 6162 GGGGCACTGCTCGCTCGCTCGAGACCGGCAACCGGAAAGAACTCAAGATTAACCTTTTC 6221
QY 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrlHiSsLysSerSerTyrlSerProGluG 2052

Db 6222 CATCCAGAACTGGAACCTCGTCTGAGTTACACCGGACAGCTTACAGCCCGCAAG 6281
QY 2052 yValGluProValSerProValSerSerProSerLeuThriAspLysGlyLeuProLy 2072
Db 6282 GGTGAGGCCCTGACGCTCTGTAGCTCACCCAGTGCAGCCACGACAGAGGGCTCCCA 6341
QY 2072 sHisLeuGluGluLeuAspLysSerThiSleuGluGluLeuArgProLysGlnProG 2092
Db 6342 GCACTGGAAGAGCTGCACAAAGCCACTGAGAGGGGAGCTGCGGCCCAACAGCAGCA 6401
QY 2092 yProValLysSleuGlyGlyGluAlaAlaHiSleuPProHisSleuArgProLeuProL 2112
Db 6402 CCGCGTGAAGCTTGGGAGAGCGCGCCCACTCCACACTGGCGCCCTGCTGAGAG 6461
QY 2112 rGlnProSerSerSerProLeuLeuGlnThraAlaProGlyValLysGlyHiSsGlnArgVa 2132
Db 6462 CAGGCCCTGTCTCAGCGCCGCTGCTCCAGACCCCGCGGGGTCAAAAGTTCACACGCGGT 6521
QY 2132 yValThrLeuAlaGlnHiSsSerGlyValIleThrGlnAspTyrlThrArgHiSsI 2152
Db 6522 GGTCAACCTTGGCCAGCATCATGAGGTATCAACAGACTACACCCGCGCACACCC 6581
QY 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrlSerPheProGlyValaSerCyAsP 2172
Db 6582 ACAGAGCTTCAGCGCACCTCTCGCGCCCGCCCTCTACTCTTCTCGGGGCAAGCTGCC 6641
QY 2172 oValLeuAspLeuAlaArgArgProProSerAspLeuTyrlleuProProProAspHiSsGlyAl 2192
Db 6642 CTTCTCGACCTTCGCGCGCCCAACCAAGTACCTTACCTCCGCGCGCGGACCATGTGTC 6701
QY 2192 aProAlaArgGlySerProHisSserGluGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCGGGCCGCTGCTCCCGCCCAAGCGAGGGGCAAGAGTCTCCAGACCAACAAAGAC 6761
QY 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetH 2232
Db 6762 GTCTGTGGGTGGGTGGAGAGCGATTTAACTGTGTCCCAACCGAGGAGCATGAC 6821
QY 2232 rGluProGlyHiSsArgSerAlaValTyrlProLeuLeuTyrlArgAspGlyGluGlnTh 2252
Db 6822 GAGCGCAGGCACTCCCGAGTGTGTGTACCGCTGTACCGGAGTGGGGAACAGAC 6881
QY 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheP 2272
Db 6882 GAGCGCCACAGATGGCTCCAGATCTCCAGGCAACACAGCCAGCCCTTCTT 6941
QY 2272 sSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLy 2292
Db 6942 CAGCAAGCTGACCGAGAGCACTCGCGCATGGTCAAGTCAAGAAAGCAAGATCAACA 7001
QY 2292 sLysLeuAsnThrHiSsAsnArgAsnGluProGluTyrlAsnIleSerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCCACCAACCGGAATGAGCTGAATACATACAGCAGCTGAGAGGA 7061
QY 2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrlArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCGCGCATACCGGAACAGGCTTATGACTTATGAAGACAGCGCGT 7121
QY 2332 lGlnGlnHiSsAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMetG 2352
Db 7122 GAGAGAACATGCGCAGACCAACATGGGCTGAGGCGCAATTAAGAAAGCACTAGGG 7181
QY 2352 yLysTyrlAspGlnTyrlGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGTGGGAAGATCCCCGCGCTCAGCGCCAAATGCTTTAACTCTGAA 7241
QY 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThraAlaAspGlyArgSerAspHi 2392
Db 7242 TGCAGTGCAGCTGCGCTGCTATGCCATTAACCGCTGTGACGAGCAGGAGTGAACA 7301
QY 2392 sThrLeuThrSerProGlyGlyGlyGlyValAlaLysValLysGlyValArgProSerSerAr 2412

Dh 7302 CACACTCACTCCAGGTGGCGGGAGAGCCCAAGGCTCTGCGACAGCCAGACCGG 7361
Qy 2412 GLYEALALYSERPROALAPROGLYLEUALASERGLYASPARPROPSERVALSERSE 2432
Dh 7362 AAAAGCCAAAGTCCCGGCGCGGCTGCGATCTGGGAGCCGGCAACCTCTGCTCTTC 7421
Qy 2432 TVALHISERGLUGLYASPCYASAAARGARGTHRPROLEUTHRANARGVALITPGIUA 2452
Dh 7422 AGTGCACTCGGAGGAGACTCAACCGCGGACCGGCTCAACACCGCTGTGGAGGA 7481
Qy 2452 PARGPRESERSEALAGLYSERTHRPROPHETRYRANPROLEULEMETAGLEUGI 2472
Dh 7482 CAGGCGCTCGTCCGACAGTTCACGCAATCCCTTACAAACCCCTGATCATGCGGTGCA 7541
Qy 2472 PALAGLYVALMETALASERPROPROPROGLYLEUPROALAGLYSERGLYPROLEUA 2492
Dh 7542 GCGCGGTCTCATGGCTTCCCAACCCCGGCTCCCGGCGGACGCGGCGGCTCTGC 7601
Qy 2492 AGLYPROHISERHISALATRAPAPGLUGIUPROLYAPROLEULEUCYSERGINTRYG 2512
Dh 7602 TGGCCCCCAACCCGCTGGAGAGAGAGCCCAAGCCACTGCTGTGCTGAGTACGAGAC 7661
Qy 2512 RLEUSERASPSERGLU 2517
Dh 7662 ACTCTCCGACAGCGAG 7677
RESULT 11
ADL13811
ID ADL13811 standard; DNA, 8686 BP.
XX
AC ADL13811;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #343.
XX
KM de; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
XX joint space narrowing; osteophyte development; joint pain;
XX osteoarthritis; SNP; single nucleotide polymorphism.
XX Homo sapiens.
XX MO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002MO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schaefer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 343; 297bp; English.
XX
XX The invention relates to a method of determining susceptibility of an
XX individual to joint space narrowing and/or osteophyte development and/or
XX joint pain comprising identifying whether the individual has at least one
XX polymorphism in a polynucleotide encoding at least one of the protein
XX listed in the specification. The method, composition and agent are
XX useful for modulating the susceptibility of an individual to joint space
XX narrowing and/or osteophyte development and/or joint pain that is
XX associated with a disease, preferably osteoarthritis. The cell line and
XX the non-human animal are useful for screening for an agent for diagnosing
XX an individual having susceptibility to joint space narrowing and/or

CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publ/published_pat_sequences).
XX
SQ Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3
US-09-522-753-5 (1-2517) x ADL13811 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Dh 157 ATGTGGGCTCCACACAGCCTGTGGACACAGCGTGAAGGCGCACTGAGCCCGCTACCGG 216
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Dh 217 CCCACAGCCTTCTTACCCAGTGCAGATGCGCCGAGCAGACAGCGAGCTGGGCTCTG 276
Qy 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Dh 277 GAGTACCAAGACCACTCCGCACTATGCTCCCACTGTGGCCCGCTCATATCAG 336
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyIleAsnGluArgSerGln 80
Dh 337 CCCAGGGGGAGAGGCCCTCCCTGCTGTGAGTTCAGCGCCGGAATGAAGCTCCAG 396
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleSerGluMet 100
Dh 397 GAGCTCACCTGGCGGCGCAGAGTCCCACTACCTGCGCGAAGTCAAGATG 456
Qy 101 GluPheIleGluSerIleArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Dh 457 GAGTCAATTAAGCAAGCCGCTCGCTAGAGCTGTGCTGACCCCTGCTGCGACG 516
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140
Dh 517 TCAACCCCTGTGGCGACGGGCGCAGCCTGCGGATGTGAACCTCAACCAAGACCGTAGC 576
Qy 141 LeuThrGlyIleLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Dh 577 CTGACGGGCAAGCTGGAAACGGGTGTCTCCCGCAGCCCGCGACACTGACCTGAGCTG 636
Qy 161 GluLeuValProProArgLeuSerIleGluIleGlnIleAsnMetAspArgValAsp 180
Dh 637 GAGCTGGTGGCGGCGGCTGTCCAAGAGAACTGATCCAAACATGAACCGCGTGGAC 696
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerIleLeuIleValIleGlnIleGlnLeu 200
Dh 697 CGAGAGATCACTCATGTGAAGAGCAGAGATCTTAAGTGAAGAAGACGAAACAGCTG 756
Qy 201 GluGluGluAlaAlaIleProProGluProGluIleProValSerProProIleGlu 220
Dh 757 GAGGAGGAGGCTCCAAAGCCGCGGAGCTGAAGAACCCCTGTACCGCGCCCATCGAG 816
Qy 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValAlaGluAla 240
Dh 817 TCAAGACCGGACCGCTGGGCGATCATATCAAGAAACGGAAGAGGCTGAAGCT 876
Qy 241 AlaHisArgIleLeuGluGluIleGluGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
Dh 877 GCAATCGAATTTGGAAGGCTGGGGCCCAAGTGAAGCTGCGGTATCAACAGGCC 936
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleValIleAsnGlnAlaMetArgIleValLeu 280
Dh 937 TCCGACACCGCGGAGTATCATGAAACATCAAAATAAACAGAGCGATGCGGAAGAGCTA 996

OY	281	lleuEutyPhelysAryGaSaBnhiSaIlaArglySgIntPrLyBgluNlyPheCySgIn	300
Db	997	ATCTGTACTTCAAGAGAGAAATCAACCTCGGAAACAATGGAGCAGAAATTCCTGCAG	1056
OY	301	ArgTyraSPGlnLeuMetGluAlaLeuGlnuLySlySVAlaGlyGllLeuAnaSPPro	320
Db	1057	CGCTATGACCAGCTCATGAGGCTCGGAGAAAGAGGGAGCGCATGGAACAACCCC	1116
OY	321	ArgArGaSaAlaLySgluSerlySVAlaArgGlyTyTyGluLyBglInPheProGluIle	340
Db	1117	CGCGCGCGGCGCAAGAGAGCAAGGTTGCGCGATCTACGAGAAAGATTCCTCGAGATC	1176
OY	341	ArgLySgluAArgGluLeuGlnGluAryMetGlnSerArGValaGlyInaArgLySergLy	360
Db	1177	CGCAAGCAGCCCGAGCTCCAGAGCGCATGCGAG--AGGGTGGCGACCGGGGAGTGGG	1233
OY	361	LeuSerMetSerAlaAlaArgSerGlnhiSgIuValaSerGluIleIleAspGlyLeuSer	380
Db	1234	CTGTCCATGTGCCCCCGCCGAGCGAGACAGAGGTGTGAGATCATGATGAGCTCTCA	1293
OY	381	GluGlnGluAnaLeuGluLySgluMetArGlnLeuAlaValaIleProProMetLeuTy	400
Db	1294	GAGCAGAGAAACCTTGAGAGAGAGATGGCCAGCTGGCGATGCCGCCATGCTGAC	1353
OY	401	AspAlaAspGlnGlnaArgIleLyPhePheIleAsnMetAsnGlyLeuMetAlaAspProMet	420
Db	1354	GACCGTGCACGACCGCATCAAGTTCATCAACAGAACGGGCTTATGGCCGACCCCATG	1413
OY	421	LySVAlaTyTLyAsPaArgInValaMetAsnMetTPSerGlnGlnGluLySgIuThPhe	440
Db	1414	AAAGTGTCAAAGACCCCGCAGGTCAATGACATGTGATGACGAGAAAGAGACCTTC	1473
OY	441	ArgGluLyPheMetGlnhiSProLySaSPheGlyLeuIleAlaSerPheLeuGluAry	460
Db	1474	CGGAGGAAGTTCATGCACATCCCAAGAACTTTGGCTGTGATGCATCATTCCTGGAGAG	1533
OY	461	LySthrValaIlaGluCyValLeuTyTyTyTyLeuThrLybLyAsnGluAnaSlyLyS	480
Db	1534	AAGACAGGCGTGAAGTGGCTCTCTATTATCACTGACTTAAGAAAGAAACACTTAAG	1593
OY	481	SerLeuValaArgArSerSerTyArArgArgArgGlyLySerSerGlnGlnGlnGlnGln	500
Db	1594	AGCTCGTGAACGAGAGCTATCGGCGCCGCGCAAGACCGACGACCAACACAGCAGAG	1653
OY	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArGSerSerGlnGluGluLyS	520
Db	1654	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCACGCCGACGACGACGAGAGAGAA	1713
OY	521	AspGluLySgluLySgluLyBgluAlaGluLySgluGluLySgProGluValaGluAn	540
Db	1714	GATGAGAAAGAGAAAGAAAGAGCGGAGAAAGAGAGAGAGAACCCGAGAGTGAAGAAC	1773
OY	541	AspLySgluAspLeuLeuLySgluLySthrAspAspThrSerGlyGluAspAsnAspGlu	560
Db	1774	GACAAAGAAAGACTCTCTCAAGAGAAAGACAGCAACCTCAGGGAGAGACACGACGAG	1833
OY	561	LySgluAlaValaAlaSerLySgluYArgLySthralaAsnSerGlnlyYArgArgLySglu	580
Db	1834	AAGAGAGCTGTGGCTCCCAAAGGCGCGCAAACTGCCCAACAGCAGGAGGAAGCGCAAAGC	1893
OY	581	ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnIn	600
Db	1894	CGCATCACCCGCTCAATGATGATAGGCAACACGAGAGAGGCCATACCCCCACGAG	1953
OY	601	SerAlaGluLeuAlaSerMetGluLeuAnaGlnuSerSerArGlyTPThrGluGluGluMet	620
Db	1954	AGCGCCAGAGCTGCGCTCATGAGAGCTGAATGAGATTTCTCGTGCACAGAAAGAAATG	2013
OY	621	GluThrAlaLybLySgluLeuGlnuIhiSgIyArGAsnTPSerAlaIleAlaArgMet	640
Db	2014	CAAAACAGCAAGAAAGTCTCTCGGAAACACGGCGCAACTGTTGGCGCATCGCCGAGATG	2073

[illegible]

Db 3102 GCCACCCCAAACTGACGCGGAGACGAGCCCTCAGACGCTGACAGACCCCG 3161
Qy 1020 GGLYLSerArGerProAlaProProAlaAspLyGlu----- 1033
Db 3162 GGGCAAGACAGAGACCCGCGACCCCGCGACAGAGAGCAGAGAACCTGTCTT 3221
Qy 1034 ----AlaPheAlaAlaGluAlaGluYLeuProGlyAspProProCysTrpThrSerG 1052
Db 3222 CCCAGCCTTGACAGCCGAGGCGCAGAAAGCTGCTGGGGAACCCCTTGCTGACCTCGG 3281
Qy 1052 YLeuProPheProValProProAlaGluValIleuYAlaSerProhiBAlaProAspPr 1072
Db 3282 CCGGCCCTTCCCGCTGCGCCCGCGTAGTGATCAAGGCTCCCGGATGCGCGAGCC 3341
Qy 1072 oSerAlaPheSerTyraLAProProGlyYhiAProleuProleuGlyLeuYhiAspThrAl 1092
Db 3342 CTCAGCCTTCTCTTAAGCTCACTGATCAGCTGACCTGCTGCGCTCCATGACCTGC 3401
Qy 1092 aATgProValLeuProArGpProProThrlIeSerAsnProProProleuIleSerSerAl 1112
Db 3402 CCGGCCCGCTGCTGCGCGCCACCACTCTCAACCCGCTCCCTCATCTCTCTGC 3461
Qy 1112 aLYhISerProSerValLeuGluuArgIuIleGlyAlaIleSerGlnGlyMetSerValG 1132
Db 3462 CAAGCACCCAGAGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGTCGTCCA 3521
Qy 1132 nLeuHISValProThySerGluYhiBAlaYAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTTCACGTCTCCAGCTCAAGACATGCCAAGCCCGCGGGCCCTTCACTCAAGGGGCT 3581
Qy 1152 uProLeuProMetAspProLySlySLeuAlaProPheSerGlyValYsgInGluInle 1172
Db 3582 GCCCGTGCATGGACCCCAAAAGCTGGCACCTTACGCGAGTAGAGAGAGAGAGCT 3641
Qy 1172 uSerProArGlyGluAlaGlyProProGlySerLeuGlyValProThralaGlnGluAl 1192
Db 3642 GTCGCCACGGGGCGAGGCTGGCGCACCGAGAGCTGGGGGTGCCACACCCAGAGAGC 3701
Qy 1192 aSerValLeuArGlyYThralaLeuGlySerValProGlyGlySerIleThrLySgYll 1212
Db 3702 GTCCGAGCTAGAGAGACAGCTCTGGGCTCAAGTTCGCGGAGAACATCCAAAGGCAT 3761
Qy 1212 eProSerThArGValProSerAspSerAlaIleThrYArGlySerIleThrHISG 1232
Db 3762 TCCAGACACGGGCTGCTCGAGAGCGCATACATACCGCGCTCCATCACCCAG 3821
Qy 1232 YThrProAlaAspValLeuTyryLySgYThrlIeThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CAGGCCAGCTGACGTCTGTAAAGGGCACATCACAGGATCATCGGAGAGAGAGCC 3881
Qy 1252 oSerArGlyLeuAspArGlyYArGlyuAspSerLeuProLySgYhiBAlaIleTyryGlu 1272
Db 3882 GAATCGCTTGAGCGCGCGGAGAGACAGCTGCCCAAGGGCCACTCACTCAAGAG 3941
Qy 1272 YLYsLYSgLYhiSValLeuSerTyryGluGlyMeSerValThrGlnCysSerLySg 1292
Db 3942 CAAGAAAGGCGCACTGTCTGTCTATAGGGTGGATGTCTGTACCAAGTCTCCAGAG 4001
Qy 1292 uAspGlyArGSerSerSerGlyProProhiSgIuThralaAlaProLyArGThrTyraS 1312
Db 4002 GAGCGCAGAAAGCAGCTCAGAGACCCCGCATGAGAGCGCGCCCAAGCGCATATGA 4061
Qy 1312 pMetMetGluGlyArGValGlyArGAlaIleSerSerAlaSerIleGluGlyLeuMetG 1332
Db 4062 CATGATGAGAGGCGCGCTGGAGAGAGCATCTCTCAAGCATCAAGAGTCTCATGGG 4121
Qy 1332 YArGAlaIleProProGlyuArgIiAspProhiShIeLeuLySgIuInhiShIeAr 1352
Db 4122 CCGTGCATCCCGCGAGAGCACAGCCCGCACACCTCAAGAGAGAGACACCATCCG 4181
Qy 1352 gGlySerIleThrGlnGlyIleProArGSerTyryValGluAlaGlnGlyuAspTyryLeuAr 1372

Db 4182 CCGGTCCATCACAAAGGATCCCTCGGTCTACGTGAGAGCACAGAGAGACTACCTCG 4241
Qy 1372 gArGluAlaLYsLeuLeuYsArGlyGluYThrProProProProProSerArGAs 1392
Db 4242 TCGGAGAGCCAAAGCTCTTAAGCGGAGAGGACAGCTTCGCGCCACCGCTTACAGGA 4301
Qy 1392 pLeuThrGluAlaTyryLySgThrGlnAlaLeuGlyProLeuLySgLeuYsProAlaHISG 1412
Db 4302 CTGACCGAGAGCTTACAAAGACGAGGCGCTGGGCGCCCTGAAGGTGAAGCGGCGCATGA 4361
Qy 1412 uGlyLeuValAlaThrVallySgIuAlaGlyArGSerIlehiSgIuIleProArGlyGlu 1432
Db 4362 GGGCTGTGGTCCACGGTGAAGAGGGCGCGCTTCATCAATGATCCCGCGAGGA 4421
Qy 1432 uLeuArGhiSThrProGlyuLeuProleuAlaProArGProleuLySgIuGlySerIleTh 1452
Db 4422 GCTGCGGACACGCGCGAGCTGCGCTGCGCGCGCGCGCGCTCAAGAGAGGCTTCATAC 4481
Qy 1452 rGlnGlyThrProleuLySgTyraPThrGlyAlaSerThrThrGlySerLySgYhiSAs 1472
Db 4482 GAGGGGACCCCGCTCAAGTACAGACCGCGGTCCACACAGTGGCTCCAAAAGACAGA 4541
Qy 1472 pValArGSerLeuIleGlySerProGlyArGThrPheProProValhiProleuAspVa 1492
Db 4542 CTAACGCTCCCTCATCGGACCGCCCGCGAGAGTTCACCGGTGACCCGCTGATGT 4601
Qy 1492 lMetAlaAspAlaArGAlaLeuGluuArgIaCyGlyTyryGluGlySerLeuYsSerArGPr 1512
Db 4602 GATGGCGAGAGCCCGGCGACTGGAAGCTGCTGTATGAAGAGAGCTGAAGAGCCGAGC 4661
Qy 1512 oGlyThralaSerSerSerGlyGlySerIleAlaArGlyAlaProValIleValProG 1532
Db 4662 AGGAGCGGACGAGCTCGGGGGGCTCATTTGGCGGGCGCCCGGTATTGGCTGA 4721
Qy 1532 uLeuGlyLYsProArGlySerpoleuLeuThrTyryGluAspHiSgYAlaProPheAlaG 1552
Db 4722 GCTGGGTAAAGCCCGGACAGGCCCTGACCTTAGAGAACAGGGGACACCTTGGCGG 4781
Qy 1552 YhiSLeuProArGlySerpProValThrMetArGgIuProThrProArGLeuGlnGlu 1572
Db 4782 CACCTTCCACAGAGTTCGCGCTGACATCGGAGAGCCACCGCGCTTGCAGAGAGG 4841
Qy 1572 YSerLeuSerSerSerLySgAlaSerGlnAspArGlySLeuThrSerThrProArGlyu 1592
Db 4842 CAGCTTGTGTCAGAGAGCATCCAGAGCGAAAGCTGACGTGACGCTGTGATGAT 4901
Qy 1592 eAlaLYsSerProhiSerThrValProGlyuhiShIeProhiAspProIleSerProTyry 1612
Db 4902 CGCCAAAGTCCCGCACAGACCGTGCCTCGAGCACACCCACACCCATCTCGCCATGA 4961
Qy 1612 uhiSLeuLeuArGlyValSerGlyValaAspLeuTyryArGSerhiSleProleuAlaPh 1632
Db 4962 GACCTGCTTGGGGGTGAGTGCSTGAGCTGTATCGAGGACATCCCGCTGGCTT 5021
Qy 1632 eAspProThrSerIleProArGlyIleProleuAspAlaAlaAlaAlaTyryLeuPr 1652
Db 5022 CAGCCCACTGCATACCCCGGGGATCCCTTGAGAGAGCGGCTGCTACACGCGC 5081
Qy 1652 oArGhiSLeuAlaProAsnProThrTyryProhiSLeuTyryProProTyryLeuIleArG 1672
Db 5082 CCGACACCTGGCGCCCAACCCCACTTACCGGACCTGTACCCACCTTACCTCCGCG 5141
Qy 1672 YTyryProAspThralaAlaLeuGluuAsnArgIuInThrlIleAsnAspTyryIleThrSe 1692
Db 5142 CTACCCCGACAGCGCGCGTGGAGAGCGGAGACATCATCAATGAATCATACCTC 5201
Qy 1692 rGlnGlnMetHiShIeAsnThrAlaThralaMetAlaGlnArGAlaAspMetLeuArG 1712
Db 5202 GAGACGATAGCACACAAACCGGACCGCATGGCCAGCGAGTGAATGTGAAGGG 5261
Qy 1712 YLeuSerProArGlyuSerSerLeuAlaLeuSntTyryAlaAlaGlyProArGlyIleI 1732
Db 5262 CCTTGTGCGCGGAGTCTGTGCTGCACTCAATGCTGGGGTCCCGAGGACATCAT 5321

QY		1732	eAspLeuSerGlnValProHisleuProValIleuValProProThrProGlyThrProAl	1752
Db		5322	GCACTGTGCCAAATGGCCACACTGCCTGTGTCTGTGTCCCCCGCACAGGACACCAGC	5381
QY		1752	aThrAlaMetAspArgIleuAlaTyrIleuProThrAlaProGlnProPheSerAraGhi	1772
Db		5382	CACGGCCATGGAGCCGCTTGCTTAACCTCCACCGCGCCCAGCCCTTCAGAACCCGCA	5441
QY		1772	sSerSerSerProLeuSerProGlyGlyProThrHisIleuThryrSerProThrThrse	1792
Db		5442	CAGACAGCTCCCACTCTCCCGAGAGAGTCCAACAACATTGAACAABACAAACACAGTC	5501
QY		1792	rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu	1812
Db		5502	CTCTCTCCAGAGCGGAGCCAGACCCGGATTCGAGACCGGGATCGGGAGCGGAAAA	5561
QY		1812	sSerIleuThrSerThrThrThrValGlnHisAlaProIleThrArgProGlyThrngl	1832
Db		5562	GTCATCTTCACTCACTCACACGAGGAGACAGCAACCATCTGAGAAGCTGTGTACGA	5621
QY		1832	uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerAraProAl	1852
Db		5622	GCAAGAGACGCGGACAGCGGACGAGAGGGGGGGGGGGGACAGACGCGCCCGC	5681
QY		1852	aSerHisSerHisSalahisGlnHisSerProIleSerProArgThrGlnAspAlaleuGl	1872
Db		5682	CTCCCACTCCATGGCCACACGACTCCGCCATCTCCCTCGAGACCCAGATGCGCTTCA	5741
QY		1872	nGluArgProSerValIleuHisAsnThrGlyMetLeuGlyLysIleThrAlaValGluPr	1892
Db		5742	GCAAGAGCCACGTGTGCTTCAACACAGGACATGAAGGTATCATCACCGCTGTGGAGCC	5801
QY		1892	oSerLyserProThrValIleuArgSerThrSerThrSerSerProValArgProAlalaThr	1912
Db		5802	CAGCAGGCCCA CGGTCCTGAGAGTCACTCCACCTCTCAACCGGTGCGCAGCTGCAC	5861
QY		1912	rPheProProAlaThrHisCySProlenGlyGlyThrIleuAspGlyValTyrProThrIle	1932
Db		5862	ATTTCCCACTCGACCCCACTGCGGCGGACCTCGATGGGGCTTACCTTACCTT	5921
QY		1932	uMetGluProValIleuLeuProLYSGluAlaProArgValAlaArgProGluArgProAr	1952
Db		5922	CATGGAGCCCTCTTGTCTGCTCCCAAGAGAGCCCCCGGGTCCGCGGACAGACGGCCCCG	5981
QY		1952	gAlaaspThrGlyHisAlaPheLeuAlaLyserProAlaArgSerGlyIleuGluProAl	1972
Db		5982	AGCAGACACCGGCCATGCTTCTCTGCAAGGCCCAACCGCGCTCGGGCTGAGAGCCGC	6041
QY		1972	aSerSerProSerLySGlySerGluProArgProLeuValProProValSerGlyHisAl	1992
Db		6042	CTCTCTCCCCCAGCAAAGGGCTCGAGACCCCGGCCCTTAGTGCTCTGTCTGTGCAAGC	6101
QY		1992	aThrIleAlaArgThrProAlaLyAsnLeuAlaProHisHisAlaSerProAspProPr	2012
Db		6102	CACCATCGCCGACCCCTCGAGAGAACTTGCACTTACCAACGCCAGCCCGGACCGGC	6161
QY		2012	oAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLyserPropheSe	2032
Db		6162	GCGCGCACTGCTCGGCTCGGACCCGCAACCGGAAAAAGACTCAATAAACCTTTTC	6221
QY		2032	rIleGlnGluIleuGluIleuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl	2052
Db		6222	CATCCAGAACTCGAACTCCGTTCTGTGGTTTACCAAGCGACAGCACTACAGCCCGAAG	6281
QY		2052	yValGluProValSerProValSerSerProSerLeuThrHisAspLySGlyLeuProLy	2072
Db		6282	GGTGAGCCCGTCAGCCCTGTGAGCTCACCCAGTGTGACCAACAGCAAGGGGCTCCCA	6341
QY		2072	sHisLeuGluIleuAspLySerHisIleuGluGlyIleuLeuArgProLyGlnProGl	2092
Db		6342	GCACCTGGAAAGAGCTCGAACAAAGCAACTGGAGGGGGAGGTGCGGCGCCAAAGACAGCAGG	6401

QY	2092	ypProValIlyLeuLgIyglYglUlaIaHIsIeuProHIsIeuAArgProIeuProgiUse	2112
Db	6402	CCCCGTGAAGCTTGCGCGGAGGCGCGCCACTCCACACACTGGCGCGCTGCTTGAG	6461
QY	2112	rgInProSeSerSeSerProIeuLeuGInTrrAlaProGIlVValIySgIlYHIsIglNaRva	2132
Db	6462	CCAGCCCTCGTCAGACCCGCGCTGCTCCAGACCGCCCGAGGGGTCAAAAGTCAACAGCGGGT	6521
QY	2132	IValThrIeuLaIaGlnHIsIIsIeSergIuValIIfhrGlnaSPYrThArghIsIePr	2152
Db	6522	GGTCACCCCTGGCCCGACACATCACTGAGGTCAATACACAGGACTACACCGGCGCACCC	6581
QY	2152	oGInGlnIeuSerAlaProIeuProAlaProIeuTYrSerPheProGIlYAlaSerCySPr	2172
Db	6582	ACACACAGCTCGCCACCCCTCGCGCCCGCCCTCTACTCTTCCCTG3GGCGAGCTGCC	6641
QY	2172	oValIeuAAspIeuArgArpProProSeSerAspIeuTYrIleuProProBAspHIsIglYAl	2192
Db	6642	CGTCTTGACCTCCGCGCGCCACCACTGACTTACTCTCCGCCCCCGGACATGGATGC	6701
QY	2192	aProAlaArgIySerProHIsIeSergIuGIlYIyIyAArgSeSerProgiUpProaenlySTH	2212
Db	6702	CCGGCCCGGTGGCTCCCCCCACAGCGAAGGGGCAAGAGGCTCTCAAGAGCCAAACAAGAC	6761
QY	2212	rSeValIleuLgIyglYglYgluApgIYIleGInProValISeSerProgiUgiYmeaTh	2232
Db	6762	GTCGGCTCTGGGTGGTGGTAGAGCGGTATTTGAACCTGTGTCCCCACCGGAGGGCAATGAC	6821
QY	2232	rgIuProgiYHIsSeSerArSeSerAlaValTYrProIeuIeuTYrArGaSPgIyGInTh	2252
Db	6822	GGAGCCAGGGCACTCCCGAGTGTGTGTACCCGCTGTGTACCGGGATGGGAAACAGAC	6881
QY	2252	rGIuProSeArArGmeGlySerIySseProgiYAsnThSeSergInProProAlaPhePh	2272
Db	6882	GGAGCCACGAGAGATGGGCTCAAGTCTCCAGCAACACGACGCGCGCACGCTTCTT	6941
QY	2272	eSeRlyLeuThrgIuSerAenSerAlaMeValIySeRtyeIySgInGInIleAenly	2292
Db	6942	CAGCAAGCTGACCGAGAGCACTCCGCATGTCAAGTCMAAGTCMAAGAGCAAGATCAACAA	7001
QY	2292	AlYleuAasnThrHIsAenArGaAnGInuProGIUTYrAsnIIsSeSergInProGIYThrgI	2312
Db	7002	GAGGTGAACACCCACAAACCGGAATGAGCTGAAATCAATTCAGCCAGCTGGGAGCGA	7061
QY	2312	uIlePheAenMeProAlaIleThrgIYThGIlYleuMeThrTYrArGSeSergInAlaVa	2332
Db	7062	GATCTTCAATATGCGCGGCATCACCGGAACAGGCTTATGACTATGAACCAACGAGCGGT	7121
QY	2332	IgInGlnHIsAlaSeThrAenMeGlyIleuGlnAlaIleIleArGlyAlaIeAenMeGI	2352
Db	7122	GCAGGAATATCCAGCACCAACATGGGGCTTGAGGCCCAATATTGAAAGGACATCATGGG	7181
QY	2352	YlyeTYrArSpGIINTpGIuGIuSerProProIeuSerAlaAenAlaPheAenProIeuAs	2372
Db	7182	TAAATATACCAAGTGGAAGAGTCCCGCGCTCAAGCCCAATGCTTTTAAACCTCTGAA	7241
QY	2372	nAlaSeRAlaSeRleuProAlaAlaMeProIleThrAlaAlaSPgIyArGSeRAspHIs	2392
Db	7242	TGCCAGTCCAGCCGCGCGCTGTATGCCATAACCGCTGTGACGGAGCGAGTAGACA	7301
QY	2392	sThrIeuThSeRProGIyglYglYglYIyAlaIyAlaIySeRGIyAspArpProProSeValISeSe	2412
Db	7302	CACACTCACTCGCCAGGTGGCGGGGGAAGCCCAAGGTCTCTGGAGACCCACACAGCGG	7361
QY	2412	gIyAlaIySeSerProAlaProGIYleuAlaISeRGIyAspArpProProSeValISeSe	2432
Db	7362	AAAACCCAAAGTCCCGCGCGCGGCTGTGCATCTGGGGACCGGCAACCTCTGTCTCTC	7422
QY	2432	rValHIsSeRgIuGIyAspCyAsnArArGThrProIeuThrAsnArGValITripIuAs	2452
Db	7422	AGTGCACTCGAGGGAGACTGCAACCGCGGACCGCGCTTACCAACCGCGTGTGGAGGA	7481
QY	2452	pArgProSeSerAlaGIySerThrProPheProTYrAsnProIeuIleMeArGLeuGI	2472

|||||
Db 7482 CAGGCCCTCGTCCAGATTCACAGCCATTCCTCCACAAACCCCTGATCATCGCGCTGCA 7541
Qy 2472 nAAGlYvAlMeAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GCGCGGTGTCATGAGCTTCCACCCCAACCGGCTCTCCCGGGGCGGCGGCTCTGC 7601
Qy 2492 aglyProHishisAlaTrpAspGluGluProLyProLeuLeuCysserGlnTyrgluTh 2512
Db 7602 TGGCCCCCACCAGCCCTGGGAGCGAGCCCAACCACTGTCTGTCTGCGAGTACGAGAC 7661
Qy 2512 rLeuSerAspSerGlu 2517
Db 7662 ACTCTCCAGACGCGAG 7677
RESULT 12
ADG86290
ID ADG86290 standard; cDNA; 8686 BP.
XX
AC ADG86290;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human SMRT encoding cDNA SEQ ID NO:4.
XX
KM SMRT; silencing mediator for retinoid and thyroid hormone action;
KM SMRT inhibitor; cytosolic; antiinflammatory; antiarthritic;
KM antirheumatic; antisense therapy; inflammatory disorder;
KM rheumatoid arthritis; hyperproliferative disorder; cancer; leukaemia;
KM breast cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 157..7680
FT /*lag= a
FT /product= "SMRT"
XX
PN MO2003106645-A2.
XX
XX 24-DEC-2003.
XX
PF 17-JUN-2003; 2003MO-US018923.
XX
PR 17-JUN-2002; 2002US-00174014.
XX
PA (ISIS-) ISIS PHARM INC.
PI Bennett CF, Freier SM, Dobie KW;
XX
XX WPI; 2004-082184/08.
DR P-PSDB; ADG86291.
DR GENBANK; AF125672.
XX
XX
PT Novel antisense compound targeted to nucleic acid encoding SMRT
PT (silencing mediator for retinoid and thyroid hormone action), useful for
PT treating animal having disease associated with SMRT such as cancer,
PT rheumatoid arthritis.
XX
XX
XX Example 13; SEQ ID NO 4; 260bp; English.
XX
XX The present invention describes a compound (I) 8-50 nucleobases in length
XX targeted to a nucleic acid molecule encoding SMRT (silencing mediator for
XX retinoid and thyroid hormone action), where (I) specifically hybridizes
XX with the nucleic acid molecule encoding SMRT and inhibits expression of
XX SMRT. (I) specifically hybridizes with at least 8-nucleobase portion of a
XX preferred target region on nucleic acid molecule encoding SMRT. Also
XX described is a composition (II) comprising (I) and a carrier or diluent.
XX (I) and (II) have cytostatic, antiinflammatory, antiarthritic and
XX antirheumatic activities, and can be used in antisense therapy, and as
XX SMRT expression inhibitors. (I) is useful for inhibiting the expression
XX of SMRT in cells or tissues. (I) is also useful for treating an animal
XX having a disease or condition associated with SMRT, e.g., inflammatory

CC disorder such as rheumatoid arthritis; or a hyperproliferative disorder
CC such as cancer chosen from leukaemia and breast cancer, by inhibiting the
CC expression of SMRT. (I) is useful for diagnostics, therapeutic,
CC prophylaxis and as research reagents and kits. The present sequence
CC encodes human SMRT, which is used in an example from the present
CC invention.
XX
SQ Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 12 Gaps: 3
US-09-522-753-5 (1-2517) x ADG86290 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyPro 20
Db 157 ATGTGGGGCTCCACACAGCCTGTGGCAGACCTGAGGGCCACTAGCCCCGCTACCCG 216
Qy 21 ProHisSerLeuSerTyProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 217 CCCACAGCCTTCTCTTACCCAGATGAGATGCCCGGACGACACAGGACGTGGGCTCTCTG 276
Qy 41 GlnTyTrpGlnHishisSerArgAspTyAlaSerHisLeuSerProGlySerIleIleGln 60
Db 277 GAGTTCACGACACACATCCCGGACTATGCTCCACCTGTGGCGGCTCATCATTCACG 336
Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 337 CCCAGCGGGAGGGCCCTCTCTGTGTGATTCAGTCCAGCCGGATGAACGCTCCAG 396
Qy 81 GlnLeuHisLeuArgProGluSerHisSerTyLeuProGluLeuGlyLeuSerGlnMet 100
Db 397 GAGCTCCACCTGGCGGACAGATGCCCATCTACCTGCGGAGCTGGGGAAGTCAGAGATG 456
Qy 101 GluPheIleGlnSerTyAspProAlaGlnLeuLeuProAspProLeuLeuArgPro 120
Db 457 GAGTTCAITGAAAGCAAGCGCCCTCGGCTAGAGCTGTGCTTACCCTGCTGGCAGCG 516
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLyAspArgSer 140
Db 517 TCACCCCTGCTGGCCACGGGCGAGCCTCGGAGATCTGAGACCTTACCAAGAGACCGTAGC 576
Qy 141 LeuThrGlyLeuLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCAGGCCCGCCGCACTGACCTGAGCTG 636
Qy 161 GlnLeuValProProArgLeuSerTyGlnGlnLeuIleGlnAsnMetAspArgValAsp 180
Db 637 GAGCTGTGGCGGCAAGCTGTCTCAAGGAGACTGATCCAGAACTGAGCCCGGTGAC 696
Qy 181 ArgGlnIleThrMetValGlnGlnGlnIleSerTyLeuLeuTyValGlnGlnGlnLeu 200
Db 697 CGAGAGTTCACCATGAGAGAGAGAGATCTTAAAGTGAAGAAAGAGCAACAGCTG 756
Qy 201 GlnGlnGlnValAlaAlaLysProProGluProGluLysProValSerProProIleGln 220
Db 757 GAGGAGAGAGCTGCCAAGCGCGCCGAGCGTGAAGAACCCGTGTACCGGCCCATCAGAG 816
Qy 221 SerTyHisArgSerLeuValGlnIleIleTyAspGlnAsnArgTyValAlaGlnVal 240
Db 817 TCCAACACCGCAGCTGTGGCAGATCATCTCAAGAGAAACCGGAAGAGCTGAAGCT 876
Qy 241 AlaHisArgIleLeuGlnGluGlyLeuGlyProGlnValGlnLeuProLeuTyTyAsnGlnPro 260
Db 877 GCACATCGATTTTGAAAGGCTGGGAGCCCGGAGTGAAGTGCCTGTACACAGGCC 936
Qy 261 SerAspThrArgGlnTyHisGlnAsnIleLysIleAsnGlnAlaMetArgTyLysLeu 280

Dh 937 TCCGACACCCGGCAGTATCATGAGAAATCAATAAACAGCGATGCGGAGAAAGACTA 996
Qy 281 ILeuTyrPheIysArgArgAsnHisAlaArgIysGlnTrpIysGlnIysPheCysGln 300
Dh 997 ATCTTGTACTTCAAGAGAGAGAAATCGCTCGGAAACAATGGAGACGAAGATTCTCCAG 1056
Qy 301 ArgTyrAspGlnIleuMetGlnIleuGlnIlyIysValGlnArgIleGlnAsnAsnPro 320
Dh 1057 CCTATGACACAGCTCATGAGAGCGCTTGGAGAAAGAAAGTGAAGCGCATCAAGAAACACCCC 1116
Qy 321 ArgArgArgAlaIlyIysIysSerIysValArgGlnTyrTyrGlnIlyIysIlePheProGlnIle 340
Dh 1117 CCGCGCGCGGCCCAAGAGAGCAAGGTTCCGAGTACTACAGAGAGAGATTCCCTGAGATC 1176
Qy 341 ArgIysGlnArgGlnIleuGlnIlyArgMetGlnSerArgValGlnIlyIysArgIysGln 360
Dh 1177 CCGAAGACCGCAGCTGACAGAGAGCCCATGACAG--AGGGTGGGCCAGCGGGCAGTGGG 1233
Qy 361 IeuserMetSerAlaIlaArgSerGlnIlyIysGlnValSerGlnIleIleAspGlyIleSer 380
Dh 1234 CTGTCCATGTCGCCCGCCGACAGCGAGCAAGAGTGTCAAGATCATCATGAGCTCTCA 1293
Qy 381 GlnGlnGlnIlyAsnIleuGlnIlyIysGlnMetArgGlnIleuAlaValIlePheProMetIleuTyr 400
Dh 1294 GAGCAGAGAACCTGGAGAGAGAGATGCGCCAGCTGACCGTGAATCCCGCCCATGCTGAC 1353
Qy 401 AspAlaAspGlnIlyIysArgIleIysPheIleAsnMetAsnGlyIleuMetAlaAspPheMet 420
Dh 1354 GACGCTGACACGAGCGCATCAAGTTCATCATCATGAACGGCTTATGGCCGACCCCATG 1413
Qy 421 IysValIlyIysAspArgGlnValMetAsnMetTrpSerGlnIlyIysGlnIlyIysIlyIysPhe 440
Dh 1414 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGAGTGAAGAGAGAGAAAGAGACTTC 1473
Qy 441 ArgGlnIlyIysPheMetGlnHisProIysAsnPheGlyIleuIleAsnSerPheIleuIlyIys 460
Dh 1474 CCGGAAAGCTTCATGAGACATCCCAAGAACTTTGGCTGATGCAATCATTTCTGAGAGAG 1533
Qy 461 IysThrValAlaGlnCysValIleuTyrTyrTyrIleuThrIlyIysAsnGlnIlyIys 480
Dh 1534 AAGACGTGGCTGATGCGTCTCTTATTACTTACTTACCTAAGAAAGAAAGAACTTAAG 1593
Qy 481 SerIleuValArgArgSerTyrArgArgIlyIysSerGlnIlyIysGlnIlyIysGlnIlyIys 500
Dh 1594 AGCGTGTGAGAGAGAGCTATCGCGCCGCGGCAAGAGCCAGCAGCAACAGCAGCAG 1653
Qy 501 GlnGlnGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 520
Dh 1654 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
Qy 521 AspGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 540
Dh 1714 GATGAGAAAGAGAAAGAAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
Qy 541 AspIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 560
Dh 1774 GACAG 1833
Qy 561 IysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 580
Dh 1834 AAG 1893
Qy 581 ArgIleThrArgSerMetAlaAsnGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 600
Dh 1894 CGCATCACCGCTCATATGCTTAATGAGGCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1953
Qy 601 SerAlaGlnIleuAlaSerMetGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 620
Dh 1954 AGCGCGAG 2013
Qy 621 GlnThrAlaIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 640
Dh 2014 GAAACAGCCCAAGAAAGGTCTCTTGAAACAAGCGCGCAACTGCTGCGCATGCGCCGATG 2073

Qy 641 ValGlySerIlyIysThrValSerGlnCysIlyIysAsnPheTyrPheAsnTyrIlyIysArgGln 660
Dh 2074 GTGGGCTCCAAAGACTGTGCGCAGTGTAAAGAACTTCTTCACTCAACAAAGAGAGCAG 2133
Qy 661 AsnIleuAspGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 680
Dh 2134 AACCTCGATGAGATCTTGAG 2193
Qy 681 ArgIlyIysIlyIysAlaProAlaAlaIleSerGlnIlyIysAlaIlePheProValVal 700
Dh 2194 AGGAAG 2253
Qy 701 GlnAspGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 720
Dh 2254 GAGATGAGAGAGATGAG 2313
Qy 721 AlaGlnIlyIysIleuIleAlaSerGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 740
Dh 2314 GCTGAA-----GCC 2322
Qy 741 ThrValAsnAsnSerSerAspThrGlnIlyIysIleProSerProIleThrGlnIlyIys 760
Dh 2323 ACTGTCAACACAGCTCAGACACAGAGCATCCCTCTCTCACTAGAGCGCGCAAG 2382
Qy 761 AspThrGlnIlyIysGlnIlyIysProIlyIysProAlaThrIleuGlnIlyIysGlnIlyIys 780
Dh 2383 GACACAGGGCAGAAATGGGCCCAAGCCCGCAGCTGGGCGCGAGCGGCGCACCCCA 2442
Qy 781 GlnProProThrProPro-ArgArgThrSerArgAlaProIleGlnProThrProAlaSer 800
Dh 2443 GGGCCACCCACCCACACAGAGAGAGATCCGGCC-CCCATGAGTCCACCCCGGCTTC 2501
Qy 800 GlnIlyIysThrGlnIlyIysProThrProProProAlaProProSerProSerAlaProPro 820
Dh 2502 TGAAGCACCTTAAGCCCTACGCCGCCACACAGACCCCATTTCTCTTCACTCTCTC 2561
Qy 820 OValIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 840
Dh 2562 TGTGGTCCCAAG 2621
Qy 840 IlyIysGlnIlyIysProProAlaIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 860
Dh 2622 GAG 2681
Qy 860 IlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 880
Dh 2682 GCTCGTCAAG 2741
Qy 880 GlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 900
Dh 2742 GAGAGCGCTGAGGCCAGCGGCGAGAGGCGCTCAAGGAGAGAGAGAGAGAGAGAGAG 2801
Qy 900 GlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 920
Dh 2802 CCGCAGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2861
Qy 920 IlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 940
Dh 2862 CTGCACTGAG 2921
Qy 940 OArgProSerIleuIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 960
Dh 2922 AAGGCCACAGCTCTCAACCCGAGCTGCGAGACCCCGGCGCAATGCTCAACCCAGAGAG 2981
Qy 960 OLeuAspIleuIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 980
Dh 2982 ACTGAGACTTGAAG 3041
Qy 980 IlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIysGlnIlyIys 1000
Dh 3042 AGTCAATGAGCCCGCGGAG 3101

QY 1000 oProProGlnAaLeuGlnProGlnuSerAaPaLaProGlnGlnProGlnuSerSerProAr 1020
 DB GCACCCCAAAACCTGACGCGGAGAGGAGGAGCCCTCAGACGCTGGACAGACCCCG 3161
 QY 1020 gGlyLysSerAArgSerProAlaProProAlaAaPlyGlu----- 1033
 DB GGGGAAAGAGAGGAGCCCGGACCCCGCGACAGAGAGGAGAGACCTGTGTTCTT 3221
 QY 1034 -----AlaPheAlaAlaGlnAlaGlnLysLeuProGlnYAspProProCysTrpThrSerG 1052
 DB 3222 CCGACGCTTGGACGCCAGAGGCCCAAGAGCTGCTGGGAGACCCCTTGCTGAGACTTCGG 3281
 QY 1052 YLeuProPheProVaLaProProArgGlnuValLLeuAlaSerProHIsAlaProaPPr 1072
 DB 3282 CTTGCTCCCTCCCGTGCCCGCGTGAGGTGATCAAGGCCCTCCCGCATGCCCCGGAGACC 3341
 QY 1072 oSerAlaPheSerTrpAlaProProGlnYHisAProLeuProLeuGlnYLeuHIsAaPThrAl 1092
 DB 3342 CTCAGCTTCTCTCAACCTCCACCTGGTCAACCACTGCCCCCTGGGCTCCATGACACTGC 3401
 QY 1092 aArgProValLeuProArProProTrpHIsAaPAsnProProProLeuHIsAaSerAl 1112
 DB 3402 CCGGCGGCTGCTGCGGCGCCACACCATCTCCAAACCCGCTCCCTCATCTCTCTGC 3461
 QY 1112 aLYsHIsAProSerValLeuGlnuArgGlnLLeuGlnAlaLLeuSerGlnLysMetSerValG 1132
 DB 3462 CAAGCAACCCAGAGCTCTCGAGAGGCAAAATAGTGCCATCTCCCAAGAAATGTGCTCCA 3521
 QY 1132 pLeuHIsAValProTrpSerGlnuHIsAAlaLYsAlaAProValGlyProValThrMetGlyLe 1152
 DB 3522 GCTCCAGCTCCCGTACTCAAGAGCATGCCAAGGCCCGGCTGGCTGTACCATGGGGCT 3581
 QY 1152 uProLeuProMetAaPProPlyLysLYsLeuAlaProPheSerGlnYValLYsGlnGlnuLe 1172
 DB 3582 GCGCTGCGCATGAGACCCCAAAAGCTGGCACTCTTACGGGAATGAAGAGGAGCGCT 3641
 QY 1172 uSerProArgGlnYGlnAlaGlyProProGlnuSerLeuGlnYValProThrAlaGlnGlnuAl 1192
 DB 3642 GTCCCAACGGGGCCAGGCTGGGCCACCGAGAGCTGGGGGTGCCCAACGCCAGAGAGGC 3701
 QY 1192 aSerValLeuArgGlnYThrAlaLeuGlnYSerValProGlnYGlySerLLeuThrLYsGlyL 1212
 DB 3702 GTCCGCTGCTGAGAGGAGACACTCGGGCTCAGTCCGGCGAAGGATCAACAAAGGAT 3761
 QY 1212 eProSerThrArgValAProSerAaPAsnAlaLLeuThrTrpArgGlnYSerLLeuThrHIsG 1232
 DB 3762 TCCAGACACAGGGGTGCTCGGACAGCGCATCACTACCGGCTCCATCACCCACGG 3821
 QY 1232 YThrProAlaAaPValLeuTrpLYsGlnYThrLLeuThrArgLLeuLLeuGlnuAaPAsnPr 1252
 DB 3822 CAGGCCAGCTGAGCTCTGTACAGAGGACCATCAACAGATCATCGGCGAGAGAGACGCC 3881
 QY 1252 oSerAArgLeuAaPArgGlnYArgGlnuAaPAsnSerLeuProLYsGlnYHisValLLeuTrpGln 1272
 DB 3882 GAGTGGCTTGACCGCGCGCGGAGGACAGCTGCGCAAGGGCCAGCTCATCAAGAGG 3941
 QY 1272 YLYsLYsGlnYHisValLeuSerTrpGlnGlnYGlyLYsMetSerValThrGlnCYsSerLYsG 1292
 DB 3942 CAAGAAAGGCGACGCTTGTCTATGAGGTGGCATCTGTGTGCCAGAGCTCTCCAGAG 4001
 QY 1292 uAaPArgLYsSerSerSerGlyProProHIsGlnuThrAlaAlaProLYsAaPTrpThrTrpAs 1312
 DB 4002 GAGCGGAGAAAGAGCTCAAGACCCCGCATGAGACGGCGCGCCCAAGGCGACCTATGA 4061
 QY 1312 pMetMetGlnGlnYArgValGlnYArgAlaLLeuSerSerAlaSerLLeuGlnLYsLeuMetG 1332
 DB 4062 CATGATGAGAGCGCGCTGGGAGAGCCATCTCTCAAGCCAGCATCAAGAGTGTCAATGG 4121
 QY 1332 YArgAlaLLeuProProGlnuArgHIsAaPProHIsAaLeuLYsGlnGlnuHIsAaLLeuAr 1352
 DB 4122 CCGTGCCATCCCGCGAGGAGACACAGCCCGCACCTCCAAAGAGAGACCAACATCCG 4181
 QY 1352 gGlySerLLeuThrGlnGlnLYsLeuProArgSerTrpValGlnuAlaGlnGlnuAaPTrpLYsLeuAr 1372

DB 4182 CGGCTCATGACACAAAGGATCCCTCGGTCTCATGTGAGGACACAGAGGATCACTGCG 4241
 QY 1372 gArgGlnuAlaLYsLeuLeuLYsAArgGlnuGlnYThrProProProProProAaPAsnArgAs 1392
 DB 4242 TCGGAGGCGAAGCTCTTAAGCGGAGGAGACCTTCGCGCCACCGCCCTCAAGGGA 4301
 QY 1392 pLeuThrGlnuAlaTrpLYsThrGlnuAlaLeuGlnYProLeuLYsLeuLYsProAlaHIsG 1412
 DB 4302 CTGACCCGAGGCTTCAAGACAGACAGGCTTGAGGCCCTTGAAGCTGAAGCGGCCATGA 4361
 QY 1412 uGlnLYsLeuValAlaThrValLYsGlnuAlaGlnYArgSerLLeuHIsGlnuLLeuProArgGln 1432
 DB 4362 GGGCTGTGGTGGCCACGGTGAAGAGAGGGGCGCCCTCATCATGATGATCCCGCCGAGGA 4421
 QY 1432 uLeuArgHIsAProGlnuLeuProLeuAlaAProArgProLeuLYsGlnLYsSerLLeuTh 1452
 DB 4422 GCTGGGAGACCGCCCAAGCTGCGCTGGCGCGCGCGCTCAAGAGAGGAGCTCATCAC 4481
 QY 1452 rGlnGlnYThrProLeuLYsTrpAaPTrpArgLYsAaSerThrGlnYSerLYsLYsHIsAs 1472
 DB 4482 GAGAGGACACCCGCTCAAGTACAGACCGGCGGTCCACCATGCTCCAAAAGACGA 4541
 QY 1472 pValArgSerLeuLLeuLYsAProGlnYArgThrPheProProValHIsAProLeuAaPAs 1492
 DB 4542 CGTAGCTCCCTCATGGGAGCCCGCGCGAGCTTCCACCGTGACCCGCTGATGT 4601
 QY 1492 pMetAlaAaPAlaArgAlaLeuGlnuArgAlaCYsTrpGlnGlnuSerLYsAaPPr 1512
 DB 4602 GATGGCGGAGCCCGGCGACTGGAACTGCTCTCTACAGAGAGCTGAAGACCCGGCC 4661
 QY 1512 oGlnYThrAlaSerSerSerGlnYGlySerLLeuAlaArgGlnYAlaProValLLeuAlProG 1532
 DB 4662 AGGAGCGGAGAGCTCGGAGGAGCTCATTTGGCGGCGCGCCCGCTCATTTGTGCTTGA 4721
 QY 1532 uLeuGlnYLYsProArgGlnuSerProLeuThrTrpGlnuAaPAsnGlnYAlaProPheAlaG 1552
 DB 4722 GCTGGGTAAACCGCGGAGAGCCCGCTGACTTATGAGACCAAGGCGACCCCTTTGCCG 4781
 QY 1552 YHisLeuProArgGlnYSerProValAThrMetAArgGlnuProThrProArgLeuGlnGlnu 1572
 DB 4782 CCACTTCCCAAGAGTTCGCGCGGTGACCATCGGAGGCCACCGCGCTGCGAGAGG 4841
 QY 1572 YSerLeuSerSerSerLYsAaPArgLYsAaPArgLYsLeuThrSerThrProArgGlnu 1592
 DB 4842 CAGCTTGTCCAGAGGAGCATCCAGACCCAAAGCTGACGTGAGCTGTGAT 4901
 QY 1592 eAlaLYsSerProHIsAaSerThrValAProGlnuHIsAaPProHIsAaPProLYsG 1612
 DB 4902 CGCGAAGTCCCGCACAGACACGCTGCCGAGACACACCCATCTCGGCTATGA 4961
 QY 1612 uHIsLeuLeuArgGlnYValSerGlnYValAaPLeuTrpArgSerHIsLLeuProLeuAlaP 1632
 DB 4962 GCACTGCTTCGGGGGTGAGTGGCTGAGCTGTATCCGAGCACATCCCTCGGCTT 5021
 QY 1632 eAaPProThrSerLLeuProArgGlnYLYsLeuAaPAlaAlaAlaLLeuTrpLYsLeuP 1652
 DB 5022 CGACCCCATCTCATACCCCGCGGACATCTCTGAGAGCGCGCTGCTCATCACTGCGC 5081
 QY 1652 oArgHIsLeuAlaAProAaPProThrTrpProHIsLeuTrpProProLYsLeuArgG 1672
 DB 5082 CCGACACCTGGCCCCCAACCCCATCAACCGGACCTGTACCCACCTCATCCGCGG 5141
 QY 1672 YTrpProAaPThrAlaAlaLeuGlnuAaPArgGlnuThrLLeuLLeuAaPTrpLYsLeu 1692
 DB 5142 CTACCCCGACACGCGCGCTGAGAGAACCGGAGACCATCATATGATCACTACCTC 5201
 QY 1692 rGlnGlnuMetHIsAaPThrAlaThrAlaMetAlaGlnuArgAlaAaPMetLeuArgG 1712
 DB 5202 GAGAGAGATGACACACACGCGGACCGCATGCGCCAGCGAGTATGCTGAAGGG 5261
 QY 1712 YLeuSerProArgGlnuSerSerLeuAlaLeuAaPTrpAlaAlaGlnYProArgGlnYLLeu 1732

Db 5262 CCTCTGCCCCGGAGTCTCTGCGACTCAACGCTGGGGTCCCCGAGGCATCAT 5321
Qy 1732 eAspleuSerGlnValProHisleuProValIleuValProProThrProGlyThrProAl 1752
Db 5322 CCACTGTCCCAAGTCCACACTGCTGTGCTGTGCCCCGACACAGGACCCACAG 5381
Qy 1752 aThrAlaMetAspArgLeuAlaTyrIleuProThrAlaProGlnProPheSerArgHis 1772
Db 5382 CACCGCATGAGACCGCTTGCTTCACTCCCAACCGGGCCCCAGCCCTTCAGACGGCCCA 5441
Qy 1772 sSerSerSerProIleuSerProGlyIleProThrHisleuThrTyrProThrThrIle 1792
Db 5442 CAGCACTCTCCCACTCTCCCAAGAGTCCACACACTTGACAAACCAACCAACGCTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu 1812
Db 5502 CTCTGTCCGAGCGGAGACGAGACCGGATCGAGACCGGAGATCGGAGCGGAGAAA 5561
Qy 1812 sSerIleleuThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrG 1832
Db 5562 GTCCATCTCTCACTCCACACGAGGTGAGACGACCCATCTGGAGACCTGTGACAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyIleGlyIleGlySerSerSerArgProAl 1852
Db 5622 GCAAGACAGCGGAGAGAGCGGACAGCGGCGGGGGGAGAGAGAGAGAGAGAGAGAG 5681
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaIleuG 1872
Db 5682 CTCCCACTCCCATGCCCACACAGCATCTGCCCATCTCCCTCGACACCGAGATCCCTCCA 5741
Qy 1872 ngInuArgProSerValIleuHisAsnThrGlyMetIleIleThrAlaValGluPro 1892
Db 5742 GCAGAGACCCAGAGTGCTGTCAACAACAGGATGAGAGGATCATCACCTGTGGAGGC 5801
Qy 1892 oSerIleProThrValIleuArgSerThrSerThrSerSerProValArgProAlaIle 1912
Db 5802 CAGCAAGCCACCGCTGTGAGGTCACCTCCACCTCTCACCCGTTGCCACCTGCGCAC 5861
Qy 1912 rPheProProAlaThrHisGlySerProIleuGlyIleThrIleuAspGlyValTyrProThrIle 1932
Db 5862 ATTCACACCTGCGACCCACCTGCCACTGGGCGGACCTCGATGGGGTCTAACCTTACCT 5921
Qy 1932 uMetGluProValIleuIleuProIleuArgIleAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGGAGCCCGTCTGTGCGCAAGAGGCCCCCGGGTGTGCGCCGAGAGCGGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheIleuAlaIleProProAlaArgSerGlyIleuGluProAl 1972
Db 5982 AGCAGACACCGGCATGCTTCTCTGCGCAAGCCCCAGCCGCTCCGGGCTGAGGCCCGC 6041
Qy 1972 aSerSerProSerIleuSerGluProArgProIleuValProProValSerGlyHisAla 1992
Db 6042 CTCTCTCCCCGAGAGGGCTCGAGAGCCCCGAGCCCTTAGTGTCTCTGTCTGTGCGCACGC 6101
Qy 1992 aThrIleAlaArgThrProAlaIleuAsnIleuAlaProHisAlaIleSerProAspProPr 2012
Db 6102 CACCATGCGCGCACCCCTCGAAGAACCTCGACCTCACACGCGGACCGCGAGCCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisArgGlyIleuThrGlnSerIleuSerIleuPhe 2032
Db 6162 GGGGCGCACCGCTGCGCTCGAGACCGCGCAACCGGAAAGACTCAAAAGTAAACCTTTTTC 6221
Qy 2032 rIleGlnIleuGluIleuArgSerIleuGlyTyrHisGlySerSerIleuSerProGlu 2052
Db 6222 CATTCAGAACTGAATCGTGTCTGTGGGTACACAGGAGAGGACTTACAGCCCGGAGG 6281
Qy 2052 yValGluProValSerProValSerSerProSerIleuThrHisAspIleuSerIleuProly 2072
Db 6282 GGTGGAGCCCGTCAAGCTGTGAGCTCACCAAGTCTGACCAAGAGAGAGGCTCCCGCAA 6341
Qy 2072 sHisIleuGluIleuAspIleuSerHisIleuGluGlyIleuIleuArgProIleuSerGlnPro 2092
Db 6342 GCACCTGGAAGAGCTGCAAGAGCCACTGAGAGGGAGAGCTGTGCGCCAAAGAGCAGG 6401

Qy 2092 yProValIleuIleuGlyIleuAlaAlaHisIleuProHisIleuArgProIleuProIleu 2112
Db 6402 CCGCGTGAAGCTTGGCGGAGAGCGCCCACTCCACACTGCGCGCCCTGTGCTAGAG 6461
Qy 2112 rGlnProSerSerSerProIleuGlnIleuThrAlaProGlyValIleuGlyHisGlnArgVa 2132
Db 6462 CAGCGCTGTGTCCAGCGCCCTGTCTCAGACCGGCGGAGGTCAAGGTCAACAGCGGGT 6521
Qy 2132 lValThrIleuAlaGlnHisIleuSerGluValIleuThrGlnAspTyrThrArgHisIleu 2152
Db 6522 GGTCACTCTGCGCCAGCATCATGATGATCATCACAGAGATTAACCGGACCAACCC 6581
Qy 2152 oGlnIleuSerAlaIleuIleuProAlaProIleuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAAGGACCGCTGCGCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6641
Qy 2172 oValIleuAspIleuArgArgProProSerSerIleuTyrIleuProProProAspHisGlyAl 2192
Db 6642 CTCTCTGAGCTTCCCGCGCCACCGAGTACCTTACCTCCGCGCCCGGACCATGATGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyIleuArgSerProGluProAsnIleuSer 2212
Db 6702 CCGGCGCCGTGTGCTCCCGCCACAGCGAAGGGGCAAGAGTCTCCAGAGCCAAACAAGAC 6761
Qy 2212 rSerValIleuGlyIleuGlyIleuAspGlyIleuGluProValSerProProGluIleuMet 2232
Db 6762 GTGCGTCTGGGTGTGTGAGGACCGGTATTGAACCTGTGTCCCAACCGGAGGCGATGAC 6821
Qy 2232 rGluProProGlyHisSerArgSerAlaValTyrProIleuIleuTyrArgAspGlyIleuGln 2252
Db 6822 GGAGCCAGGAGCATCTCCGAGTGTGTGACCGGTGTGTACCGGAGTGGGAGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerIleuSerProGluIleuThrSerGlnProProAlaPheP 2272
Db 6882 GGAGCCAGAGATGGCTCCAGATGTCTCAAGTCTCCAGGCAACACAGCAGCCGACCTTCT 6941
Qy 2272 eSerIleuSerThrGluSerAsnSerAlaMetValIleuSerIleuGlnIleuAsnIleu 2292
Db 6942 CAGCAAGTGAACCGAGAGCACTCCGCAATGTGCAAGTCCAGAGAACAGAGATCAACAA 7001
Qy 2292 sIleuIleuAsnThrHisAsnArgAsnGluProGluTyrAsnIleuSerGlnProGlyThr 2312
Db 7002 GAAGCTGAACACCCACAAACCGGATAGCCTGAAATCAATATGACGAGCTTGAGCGGA 7061
Qy 2312 uIlePheAsnMetProAlaIleuThrGlyIleuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCTCCGCCATCACCGGAACAGGCTTATGACCTTATGAAAGCCAGCGGT 7121
Qy 2332 lGlnIleuHisAlaSerThrAsnMetGlyIleuGluAlaIleIleArgIleuAlaIleuMetG 2352
Db 7122 GCAAGAACATGCCAGACCAACATGGGGCTGGAGGCCATATATGAAGGCACTATGGG 7181
Qy 2352 yIleuTyrAspGlnTTPGluIleuSerProProIleuSerAlaAsnAlaPheAsnProIleuAs 2372
Db 7182 TAAATATGACCAAGTGGGAAGAGTCCCGCGCTCAGCGCCCAAGCTTTTAACCTCTGAA 7241
Qy 2372 nIleAsnAlaSerIleuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
Db 7242 TGCAGGTGCGAGCTGCGCTGTGTATGCCATTAACCGTGTGAGCGAGAGGTATACCA 7301
Qy 2392 sThrIleuThrSerProGlyIleuGlyIleuAlaIleuValIleuSerGlyArgProSerSerAr 2412
Db 7302 CACACTCACCCTGCCAGGTGGGGGGGAAAGGCAAGGTCTCTGGAGACCCAGAGCCG 7361
Qy 2412 gIleuAlaIleuSerProAlaProGlyIleuAlaSerGlyIleuAspArgProProSerValSerSe 2432
Db 7362 AAAAGCCAAGTCCCGCGCCCGCGCTGTGATCTGGGAGACCGGCAACCCCTGTCTCTC 7421
Qy 2432 rValHisSerGluIleuIleuAspCysAsnArgArgThrProIleuThrAsnArgValTTPGluAs 2452
Db 7422 AGTGCACTGGAAGAGAGATGCAACCGCGGAGAGCGGCTTCAACACCGCGTGTGGAGGA 7481

Qy 2452 parGProSerSerA1aGlySerThrProPheProTyAanProLeuIleMetAArgLeuG1 2472
Db 7482 CAGGCCCTCGTCCGACAGTTCACGCCATTCCTCCACAAACCCCTGATCATGCGGCTGCA 7541
Qy 2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuA1 2492
Db 7542 GGCAGGCTGTCATGGCTTCCTCCACCCCAACCGGCTCTCCCGGAGCGGCGGCTCGC 7601
Qy 2492 aGlyProH1aH1a1aTTPaAPG1uG1uProTyAanProLeuIleCysSerGlyProGlyLeu 2512
Db 7602 TGGCCCCCACCACCGCTGGAGAGAGAGACCCCAAGCCACTGCTGTGCTGCACTACAGAAC 7661
Qy 2512 rLeuSerAPSerGlu 2517
Db 7662 ACTCTCCACAGCGAG 7677

RESULT 13

ADQ18920

ID ADQ18920 standard; DNA; 8686 BP.

XX AC ADQ18920;

XX DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1739.

XX KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

XX de.

XX OS Homo sapiens.

XX PN MO2004048938-A2.

XX PD 10-JUN-2004.

XX PF 26-NOV-2003; 2003WC-US038193.

XX PR 26-NOV-2002; 2002US-0429739P.

XX PA (PROT-) PROTEIN DESIGN LABS INC.

XX PI Aziz N, Ginsburg WM, Zlotnik A;

XX DR WPI; 2004-441208/41.

XX XX

PT Early detection of soft tissue sarcoma comprises determining expression
of a gene in a first soft tissue sample and a normal soft tissue sample
and comparing the gene expression, also useful in treating soft tissue
sarcoma.

PS Example 2; SEQ ID NO 1739; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.

SQ Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0
Score: 12978.00
Percent Similarity: 98.57%
Best Local Similarity: 98.46%

Length: 8686
Matches: 2487
Conservative: 3
Mismatch: 9

Query Match: 98.21% Indels: 28
DB: 12 Gaps: 3
US-09-522-753-5 (1-2517) x ADQ18920 (1-8686)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPaArgAlaThrGluProArgTyPro 20
Db 157 ATGTGGGGCTTCACACAGCCGTGTGCAACAGCGTGAAGGGCCATGAGCCCCGCTACCCG 216
Qy 21 ProH1aSerLeuSerTyProProValGln1leAlaArgThrH1aThrAPValGlyLeuLeu 40
Db 217 CCCACAGCCCTTCTCCACAGTGCATGATGCCCGGAGCAGACAGAGCTCGGCTCTCG 276
Qy 41 GluTyGlnH1aH1aSerAPTyA1SerH1aSerLeuSerProGlySer1le1leGln 60
Db 277 GAGTACAGACACCACTCCCGGACTATGCTCCACCTGTCTGCCGCTCATATCCAG 336
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAanGluArgSerGln 80
Db 337 CCCAGCGGCGAGAGCCCTCCCTGCTGTGAGTTCAGCCCGGGAATGAACGGTCCAG 396
Qy 81 GluLeuH1aSerLeuArgProGluSerH1aSerTyLeuProGluLeuGlySerGluMet 100
Db 397 GAGCTCCACCTCGGCGCAGAGTCCACTCATACCTGCCAGCTGAGAGTCAAGATG 456
Qy 101 GluPhe1leGluSerTyA1SerProArgLeuGluLeuLeuProAPProLeuLeuArgPro 120
Db 457 GAGTTCATTGAAGCAAGCCCTCGGCTAGAGCTGTGCTGACCCCTGCTGCAACG 516
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAanLeuThrTyAanArgSer 140
Db 517 TCAACCTGCTGGGCAAGCGGCGCAGCCCTGCGGATCTGAAGACCTCAACAGACCTGAGC 576
Qy 141 LeuThrGlyLeuGluProValSerProProSerProProH1aThrAPProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAAACCGGTCTCCCGGAGCCCGGACACATGACCTGAGCTG 636
Qy 161 GluLeuValProProArgLeuSerTyGluGluLeu1leGlnAanMetAPArgValAsp 180
Db 637 GAGCTGTGCGCCGACAGGCTGTCCAGAGAGAGCTATCCAGAACATGACCGCTGAGC 696
Qy 181 ArgGlu1leThrMetValGlnGlnGln1leSerTyAanLeuTyAanGlnGlnLeu 200
Db 697 CGAAGATCACCATGTGAAGAGCAGATCTTAAGCTGAAGAAAGAGCAGAACACTG 756
Qy 201 GluGluGluAlaAlaTyProProGluProGluTyProValSerProProPro1leGlu 220
Db 757 GAGGAGAGGCTGCCAAGCGCGGAGCTGAAGAGCCGCTGTACCGCGCCATGAG 816
Qy 221 SerTyAanArgSerLeuValGln1le1leTyAanProLeuArgTyAanGluAla 240
Db 817 TCAGAGCACCGGAGCTGTGTGATCATTTACAGAGAAACCGGAGAGGCTGAGAGCT 876
Qy 241 AlaH1aArg1leLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyAanGlnPro 260
Db 877 GCAATCGGATTCGGAAGGCTGGGGGCCCAAGTGAAGCTGCGCTGTAACAACAGCCC 936
Qy 261 SerAspThrArgGlnTyH1aGlnAan1leTyAanGlnAlaMetAArgTyAanLeu 280
Db 937 TCGAGACCCCGGAGATCATGTGAACATCAATAATAAACAGGCGATGCGGAGAGCTA 996
Qy 281 IleLeuTyPheTyAanArgAanH1aAlaArgTyGlnTyProGlnTyAanPheCysGln 300
Db 997 ATCTGTACTTCAGAGAGAGATTCACGCTCGAAGAACATGAGAGCAGAGATTCTGCCAG 1056
Qy 301 ArgTyAanPGLuLeuMetGluAlaLeuGluTyAanValGluValGluArg1leGluAanPro 320
Db 1057 CGCTATGACAGGCTCATGGAGGCTCGGAGAAAGTGAAGCGCATTCGAGAACACCCC 1116
Qy 321 ArgArgArgAlaAlaLeuSerTyAanArgGluTyTyGluTyAanPheProGlu1le 340
Db 1117 CGGCGGCGGCGCAAGAGAGCAAGTTCGAGATCTTACGAGAAAGATTCCTCGTGAATC 1176

QY 341 ArgLySGlnArgGluLeuEngInGlnArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db 1177 CCGAAGCACGCGAGCTGACGAGCGCATGACAG---AGGGTGGCCACGCGGCGAGTGGG 1233
QY 361 LeuSerMetSerAlaAlaArgSerGluHlaGluValSerGluIleIleLeuArgGlyLeuSer 380
Db 1234 CTGTCAATGTGCGCCCGCCGACGAGCACGAGGTGTCAAGATGATCATGATGCGCTCTCA 1293
QY 381 GlnGlnGlnLeuLeuGlnLeuArgGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
Db 1294 GACGAGGAGAACCTGGAGAAAGCATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTAC 1353
QY 401 AspAlaAspGlnGlnArgIleLeuysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
Db 1354 GACGCTGACACGACGACATCAAGTTCATCAATGAAACGGGCTTATGCGCGACCCCATG 1413
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGlnLysGluTyrPhe 440
Db 1414 AAGGTGTACAAAGACCGCCAGTCATGAACATGTGAGTGAAGACGAGAGAGACCTTC 1473
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
Db 1474 CCGGAGAAAGTTCATGACGATCCCAAGAACTTGGCTGATCCGATCATTCTCGAGAGG 1533
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysAsnGluAsnTyrLys 480
Db 1534 AAGACAGTGGCTGAGTGGCTCTCTATTACTGACTGAAGAAAGATGAACTATTAAG 1593
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1594 AGCTGTGTAGAGAGAGCTATCGCGCGCGCAAGAGCCAGACGACAAACAGCAGCAG 1653
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1654 CACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
QY 521 AspGluLysGlnLysGlnLysGlnValGluLysGlnGluLysProGluValGluAsn 540
Db 1714 GATGAAGAGAGAGAAAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
QY 541 AspLysGluAspLeuLeuLysGlnLysThrAspAspThrSerGlyLysAspAspAspGlu 560
Db 1774 GAAAG 1833
QY 561 LysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGly 580
Db 1834 AAGGAGGCTGTGCTCCAAAGCCGCAAACTGCCAAACGCCAGGAAACCCGAAAGGC 1893
QY 581 ArgIleThrArgSerMetAlaAsnGluValaAsnSerGluAlaIleThrProGlnGln 600
Db 1894 CGCATCACCGGCTCAATGGCTTAATGAGGCCAACAAGCAGAGAGGCCATCACCCCCAGCAG 1953
QY 601 SerAlaGluLeuLeuLeuSerMetGluLeuAsnGluSerSerArgTyrPheGluGluMet 620
Db 1954 AGGCGGAGGTGGCTCTCATGAGCTGAATGAGATGTTCTGTGAGACGAGAGAGAAATG 2013
QY 621 GlnThrAlaLysLysGlyLeuLeuGlnHlaGlyArgAsnTyrPheSerAlaIleAlaArgMet 640
Db 2014 GAAACAGCCAAAGAAAGTCTCTCGAACAAGCCGCAACTGTGCGCACTGCCCGGATG 2073
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysValArgGln 660
Db 2074 GTGGGCTCCAAAGCTGTGTGCGAGTGAAGAACTTCACTTCAACCTCAAGAAAGAGCAG 2133
QY 661 AsnLeuAspGluIleLeuGlnGlnHlaLysLeuLysMetGluLysGlnValGlnAlaArg 680
Db 2134 AACCTGATGATGATCTTGACAGACACAACTGAAGATGAGAGAGAGAGAGAGAGAG 2193
QY 681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGlnAlaAlaPheProProValVal 700
Db 2194 AGGAAAG 2253
QY 701 GluAspGlnGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGln 720

Db 2254 GAGGATGAG 2313
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyCylCysSerGlyProAla 740
Db 2314 GCTGAH-----GCC 2322
QY 741 ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisThrGluAlaAlaLys 760
Db 2323 ACTGTCAACAACAGCTCAGACACCGAGACATCCCTCTCTCACTGACAGGCGCCAGAG 2382
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyValaAspGlyProProPro 780
Db 2383 GACACAGGCGCAGATGGGCCCAAGCCCAAGCCACTGGCGCCGACAGGAGCCACCCCA 2442
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db 2443 GGGCCACCCACCCACACCGAGAGACATCCCGGC-CCACAGATTCACCCCGGCTC 2501
QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACTTATAGCCCTACGCGCCACACAGACCCCATTTCCCTTTCACCTCTCC 2561
QY 820 oValValProLysGlnGluLysGlnGluGluThrAlaAlaAlaProProValGluGlu 840
Db 2562 TGTGTCTCCCAAG 2621
QY 840 yGluGluGlnLysProProAlaAlaGluGluLeuAlaValaAspThrGlyLysAlaGlu 860
Db 2622 GAG 2681
QY 860 uProValLysSerGlnCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db 2682 GCGGCTCAAGAGAGAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2741
QY 880 eGluAlaAlaGluAlaThrAlaGluGluValaLeuLysAlaGluLysLysGluGlySe 900
Db 2742 GAGAGCGCTGAGAGCCACGCGCCGAGAGAGCGCTCAAGGAGAGAGAGAGAGAGAGAG 2801
QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Db 2802 CGGACAGGCGCACACAGCCAGAGAGCTCGGGCGCCCGCCAGAGAGAGAGAGAGAGAG 2861
QY 920 rCysSerAlaAspGluValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
Db 2862 CTGCACTGCAGAGAGATGATGAGCGCGAGCGCGCGCGAGCAAAACCGGCTGTCTCC 2921
QY 940 oArgProSerLeuLeuThrProThrGlyAspProAlaAlaAsnAlaSerProGlnLysPr 960
Db 2922 AAGGCCACGCTCTCACCCCGACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2981
QY 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLys 980
Db 2982 ACTGACCTGAACACACTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3041
QY 980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Db 3042 AGTCCATGAGCCCCCGGAG 3101
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlnLysSerProAr 1020
Db 3102 GCCACCGCAAAACCTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3161
QY 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3162 GGGCAAG 3221
QY 1034 ---AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGln 1052
Db 3222 CCGAGCCTTGAG 3281
QY 1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072

Db 3382 CTTGCCCTTCCCGTCCCGCCCGGAGTGATCAAGGCTTCCCGCATGCCCGGAGACC 3341
Qy 1072 oSeRAlApheserTyrAlaProProGlyVhiAProleuProleuGlyLeuH1saapThra1 1092
Db 3342 CTCAGCCTTCTCTCAAGCTTCAACCTTGACCACTGCGCTCGGCTCTCAAGACAGC 3401
Qy 1092 aArgProValLeuProArgProProThri1seSerAsnProProleu11seSerAl 1112
Db 3402 CCGGCGCGTCTGCGCGGCCACCAACCATCTCCAAACCGGCTCCCTCATCTCTGCG 3461
Qy 1112 alyvhiAProserValLeuGlyuaArgIn11eGlyAla11eSerGInGlyMeSerValG1 1132
Db 3462 CAACACCCCAAGCGCTCTCGAGAGCAAAATAGTGCCATCTCCAAAGAAAGTGCGGTCCA 3521
Qy 1132 nLeuH1eValProTyrSerGInH1sa1a1yAlaProValG1ProVal1ThrMeGlyLe 1152
Db 3522 GCTTCAGTCTCCGTACTCAAGCATGCGAAGGCCCGGTGGGCTGTGACCATGAGGCT 3581
Qy 1152 uProLeuProMeAsnProLySlyLeuAlaProPheserGlyVal1ySGInGlyInLe 1172
Db 3582 GCCCTGCGCATGAGCCCAAAAGCTGGCACTTCAAGCGAGTGAAGCAAGAGAGCT 3641
Qy 1172 uSerProArgGlyGIn1aGlyProProGlyuSerLeuGlyVal1ProThra1aGInGlyAl 1192
Db 3642 GTCCCAAGCGGCGCAGGCTGGGCGCACCGAGAGCTGGGGGTGCCCAAGCCCAAGAGGC 3701
Qy 1192 aSerValLeuArgGlyThra1aLeuGlySerValProGlyGlySer11eThrLySGly11 1212
Db 3702 GTCCGTCTGAGAGGAGCAAGCTTGAGCTCAGTTCGGGCGGAGACATCAACAAAGCAT 3761
Qy 1212 eProSerThraArgValProSerAsnSerAla11eThrTyrArgGlySer11eThrH1sG1 1232
Db 3762 TCCAGACACAGGGTGCCCTCGACAGCGCATCAATACCGCGCTCTCAACCCAGC 3821
Qy 1232 yThnProAlaAspValLeuTyrLySGlyThri1eThraArg11e11eGlyGlyuAspSerPr 1252
Db 3822 CACCCAGCTGACCTCTCTGTAAGGAGCACCATCAAGAGATCTCGCGAGAGACACCC 3881
Qy 1252 oSeRArgLeuAspArgGlyThraGlyuAspSerLeuProLyGlyVhi11eTyrGlyG1 1272
Db 3882 GAGTGGCTTGAGAGCGGCGCGGAGAGCAAGCTGCCCAAGGGCCACGTCATCAAGAG 3941
Qy 1272 yLySlyGlyVhi1eValLeuSerTyrGlyGlyGlyMeSerValThrGInCySerLySG1 1292
Db 3942 CAAGAAAGGCGCACTCTTGTCTATGAGGGTGGAGATCTGTGACCCAGTCTCAAGAG 4001
Qy 1292 uAspGlyArgSerSerSerGlyProProH1sGInThraAlaProLyAspArgThryrAs 1312
Db 4002 GGAAGGAGAGAGCTCAGGAGCCCCCATGAGACGGCGCCCAAGCGCACCTATGA 4061
Qy 1312 pMeCMeGlyGlyVArgValaGlyArgAla11eSerSerAlaSer11eGlyGlyLeuMeG1 1332
Db 4062 CATGATGAGAGGCGCGTGGGAGAGCATCTCTCAAGCCAGCATCAAGAGTCTCAAGG 4121
Qy 1332 yATGAla11eProProGlyuArgH1sSerProH1sH1eLeuLySGInGlyH1sH1s11eAr 1352
Db 4122 CCGTGCATCCCGCGGAGGACACAGCCCGCACCATCTCAAGAGAGACACCACTCG 4181
Qy 1352 gGlySer11eThrGInGly11eProArgSerTyrValaGlyuAGInGlyuAspTyrLeuAr 1372
Db 4182 CCGGTCATCAACAAGAGATCCCTCGTCTTATGAGAGGACAGAGAGCATCTCGCG 4241
Qy 1372 gATGAGlyuAlaLeuLeuLeuLyAspArgGlyGlyThrProProProProProProAspArg 1392
Db 4242 TCGGAGAGGCAAGCTCTAAAGCGGAGGAGGACGCTCCGCGCCACCGCTCAAGGGA 4301
Qy 1392 pLeuThGlyuAlaTyrTyrThrGIna1aLeuGlyProLeuLyLeuLyAspProAlaH1sG1 1412
Db 4302 CTGAGCGAGAGGCTCAAGAGAGAGGCGCTGGGCGCTCAAGTGAAGCGCGCATGA 4361
Qy 1412 uGlyLeuVala1aThraVallySGlyuAGlyArgSer11eH1sGInu11eProArgGlyG1 1432
Db 4362 GGGGCTGTGGCCAGCTGAGAGAGGCGGCGCTCATCATGATGATCCGCGCGAGGA 4421

Qy 1432 uLeuArgH1eThrProGlyuLeuProleuAlaProArgProleuLySGlySer11eTh 1452
Db 4422 GCTGGGACACAGGCGGAGAGCTGCTGGCCCCCGGCGGCTCAAGAGGCTCATATAC 4481
Qy 1452 rGInGlyThrProleuLyTyrAspThrGlyAla11eSerThrGlySerLySlyH1sAs 1472
Db 4482 GCAAGGACACCGGCTCAAGTGAAGACAGGCGGCTGACCACTGGCTCAAAAGACAGA 4541
Qy 1472 pValArgSerLeu11eGlySerProGlyVArgThrPhaProProValH1sProLeuAspVa 1492
Db 4542 GTAGGCTCTCTCATGCGAGAGCCCGGCGGAGCTTCCACCGGTGACCGCGTGAAGT 4601
Qy 1492 lMeRAlAspAlaArgAlaLeuGlyuaArgAlaCyTyrGlyGlyuSerLeuLySAspArgPr 1512
Db 4602 GATGGCGGACCGCGGCGCTGGAACGTGCTGTCAAGAGAGAGCTGAAGAGCGGCC 4661
Qy 1512 oGlyThra1aSerSerSerGlyGlySer11eAlaArgGlyAlaProVal11eVal1ProG1 1532
Db 4662 AGGAGCGCGAGAGCTCGGGGGCTCATTTGGCGCGGCGCCCGGTCAATTGTGCTGA 4721
Qy 1532 uLeuGlyLyProArgGlyuSerProleuThrTyrGlyuAspH1sGlyAlaProPha1aG1 1552
Db 4722 GCTGGTGAAGCGCGGAGAGCCCTGACCTTGAAGAGACACAGGCGCACCTTTGCGG 4781
Qy 1552 yH1sLeuProArgGlySerProVal1ThrMeArgGlyuProThrProArgLeuGInGlyG1 1572
Db 4782 CCACCTTCCACAGAGTTCGCCCTGACCATAGCGGAGCCACCGCGGCTTCAAGAGG 4841
Qy 1572 ySerLeuSerSerSerLyAlaSerGInAspArgLyLeuThrSerThrProArgGlyu1 1592
Db 4842 CAGCTTTCGTTCAGAGAGGATCCAGAGCCGAAGCTAGCGTCAAGCTCTGTAGAT 4901
Qy 1592 eAlaLySerProH1sSerThra1ProGlyuH1sH1eProH1sSerProTyrG1 1612
Db 4902 CGCAAGTCCCGGAGAGCAAGCTGGCGGAGACACACACCCATCTCGCCCTATGA 4961
Qy 1612 uH1sLeuLeuArgGlyVAlaSerGlyValaAspLeuTyrArgSerH1s11eProleuAlaPh 1632
Db 4962 GCACCTCTTCCGGGCGTGAAGCTGAGACCTGTATCGAGCCACATCCCTTGCGCTT 5011
Qy 1632 eAspProThSer11eProArgGly11eProleuAspAlaAla1aTyrTyrLeuPr 1652
Db 5022 GAGCCCATCTCATATACCCCGGCGCATCCCTGGAAGAGCGGCTCTATCACTAGCC 5081
Qy 1652 oATGHisLeuAlaProAsnProThrTyrProH1sLeuTyrProProTyrLeu11eArgG1 1672
Db 5082 CCGACACTGGCCCCCAACCCCACTACCGGACCTGTACCCACCTTACCTCATCCGCG 5141
Qy 1672 yTyrProAspThra1aAlaLeuGlyuAspArgGInThri11e11eAsnAspTyr11eThrSe 1692
Db 5142 CTACCCCGAGAGGCGGCGCTGAGAGAGCGGAGAGCATCATCATATCATCATCTTC 5201
Qy 1692 rGInGInMeTh1sH1sAnThra1aThra1aMeAlaGInaArgAlaAspMeLeuArgG1 1712
Db 5202 GAGAGAGATCAACAACAAGGCGGACCGGCAATGGCCCAAGAGAGTATGTGAAGGG 5261
Qy 1712 yLeuSerProArgGlyuSerSerLeuAlaLeuAsnTyrAla1aGlyProArgGly11e11 1732
Db 5262 CCTCTGCGCCCGGAGAGCTCTGCTGGCATCACTAGCTGGCGGCTCCCGAGGAGCAT 5321
Qy 1732 eAspLeuSerGInValaProH1sLeuProValaLeuValaProProThrProGlyThrProAl 1752
Db 5322 CGACCTTCCCAAGTGCACACCTGCTGTGCTGTGCCCCCGAGACCAAGACCCCAAGC 5381
Qy 1752 aThra1aMeAspArgLeuAlaTyrLeuProThra1aProGInProPheserSerArgH1 1772
Db 5382 CACCGCATGAGAGCGCTTCTTACCTCCCAAGCGGCGCCAGCCCTTCAAGAGCGGCA 5441
Qy 1772 aSerSerSerProleuSerProGlyGlyProThrH1sLeuThrLyAspProThrThrSe 1792
Db 5442 CAGCAGTCCCACTCTCCCAAGAGGTCMAACAACCTTGAACAAACCAACCAAGTGC 5501

1792 rSerSerGIuArGluArGAspArGAspArGluArAspArGAspArGluArGluY 1812
5502 CTGCTCCGAGCGGGAGCGAGACCGGGATGTAGAGCGGGACCGGGAGCGGGAAAA 5561
1812 sSerTlleuThrSerThrThrValGluHisAlaProIleTPArGProGlyThrG 1832
5562 GTCCAACTTCACGTCCACACGAGGTGAGCAGCACCATTGAGACCTGTGTCAAA 5621
1832 uGInSerSerGIySerSerGIySerSerGIyGlyGlyGlySerSerSerArGProAl 1852
5622 GCAAGACAGCGGAGAGCGGAGCAGCGCGGGGTTGGGGGACGACGACCGCCCGC 5681
1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArGThrGlnAspAlaLeuG 1872
5682 CTCCCACTCCCAATGCCACACGACACTGGCCACTCCCTCCGAGCCAGATCCCTCCA 5741
1872 nGlnArGProSerValLeuHisAsnThrGlyMetGlyGlyLeileThrAlaValGluP 1892
5742 GCAGAGACCCAGTGTCTTCAACAACAGGCATGAAGGGTATCATCACCGCTGTGAGCC 5801
1892 oSerLySPromThrValLeuArGSerThrSerThrSerSerProValArGProAlaIaTh 1912
5802 CAGCAAGCCCAACGCTCTGAGATCCACTTCCACTCTTCACTCCCTTCCGACCTGCC 5861
1912 rPheProProAlaThrHisCysBProLeuGlyGlyThrLeuAspGlyValIyrBProThle 1932
5862 ATTCCCACTGCACCCACTGTGGCGGACCTCGATGGGGTGTACCTTACCTTACCT 5921
1932 uMetGluPProValLeuLeuPProLyGlyAlaProAlaGValAlaArGProGluArGProAr 1952
5922 CATGGAGCCGCTTCTCTGCCCCAAGAGGCCCGCGGCTGCCGCGCAAGGGCGCCG 5981
1952 gAlaAspThrGlyHisAlaPheLeuAlaLySPromProAlaArGSerGIyLeuGluProAl 1972
5982 AGCAGACACCGGACCAAGCTTCTTCCGCAAGCCCCCAAGCCCCGCTCCGGCTGAGCCCGC 6041
1972 aSerSerProSerLyGIySerGIuProArGProLeuValProProValSerGIyHisAl 1992
6042 CTCTCTCCCGAGCAAGGGCTCGAGGCCCGGCCCTTAGTGTCTCTGTCTGCGCACGC 6101
1992 aThrTlleAlaArGThrProAlaLyAsnLeuAlaProHisHisAlaSerProAspProP 2012
6102 CACCAATCGCCGACCCCTGCGAAGAACTTCGCACTCACACGACCGCAGCCGACCGCC 6161
2012 oAlaPProAlaSerAlaSerAspProHisArGGLuLySPThrGlnSerLySPProPhe 2032
6162 GCGGCACTGCTCGGCTCGACCGCACCGGAGAAAGACTCAAGATTAACCTTTTTC 6221
2032 rIleGlnGluLeuGluLeuArGSerLeuGlyTyrHisGlySerSerTySerProGluG 2052
6222 CATCCAGGAACGTGAACTCGCTTCTGTGGGTACACAGGCAGCAGTACAGCCCGGAGG 6281
2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLySGlyLeuPProLy 2072
6282 GGTGGAGCCCGTACGCTGTGACTCACCACTGACCCACAGCAAGGGGCTTCCCAA 6341
2072 sHisLeuGluGluLeuAspLySerHisLeuGluGlyGluLeuArGProLySGlnProG 2092
6342 GACCTGGAGAGACTGACAGAGCAACCTGAGGGGAGACTGCGGCCCAAGACCGCAGG 6401
2092 yProValLyLeuGlyGlyGluAlaHisLeuProHisLeuArGProLeuPProGluSe 2112
6402 CCCCGTAGAGCTTGGCGGGAGCGCGCCACTCCCACTCGGGCGCTGCTGAGAG 6461
2112 rGlnPProSerSerSerProLeuLeuGlnThrAlaProGlyValIlySGlyHisGlnArGVA 2132
6462 CAGGCCCTGCTCCAGCCGCTGTCCAGACCGCCCAAGGGGTCAAAAGTACACAGCGGGT 6521
2132 lValThrLeuAlaGlnHisIleSerGIuValIleThrGlnAspTyThrArGHisIleSP 2152
6522 GGTCAACCTTGCCAGACATCATGTAGGTATCATCACAGGAATACACCGGACCAACCC 6581
2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCysPr 2172

6582 ACAGCAGCTCACGCGACCCCTCGCCGCCCTTCACTCTTCCCTGGGGCCAGCTGCC 6641
2172 oValLeuAspLeuArGArGProProSerAspLeuTyTleuPProProProAspHisGlyAl 2192
6642 CGTCTGAGACTTCGCGGCCCAACCACTTCACTTACCTTCCGCCCGGACCATGGTGC 6701
2192 aProAlaArGlySerProHisSerGIuGlyGlyLyAspSerProGluProAsnLyTh 2212
6702 CCGGGCCGCTGCTCCCGCCACAGCGAAGGGGCAAGAGTCTCCAGAGCCAAACAAGAC 6761
2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
6762 GTCGGTCTGGGTGGGTGAGGACCGGTATTGAACCTGTGTCCCAACCGAGGGCArGAC 6821
2232 rGluPProGlyHisSerArGSerAlaValIyrProLeuLeuTyTArGAspGlyGlyGlnTh 2252
6822 GAGGCCAGGCACTCCCGAGGTGTGTGTACCGCTGTGTACCGGAGTGGGACAGAC 6881
2252 rGluProSerArGMetGIySerLySerProGlyAsnThrSerGlnProProAlaPheP 2272
6882 GAGGCCACAGGATGGCTCCAAAGTCTCACGCAACACAGCCAGCCGCAAGCTTCTT 6941
2272 eSerLyLeuThrGluSerAsnSerAlaMetValLySerLyBlyGlnGluIleAsnLy 2292
6942 CAGCAAGCTGACCGAGACCACTCCGCAATGTCAGTCAAGTCAAGAACCAAGATCAACA 7001
2292 sLyLeuAsnThrHisAsnArGAsnGluPProGluTyTAsnIleSerGlnProGlyThrG 2312
7002 GAACTGAACACCCACACACCGAATAGCTGAATCAATATACAGCAGCTGGACGGA 7061
2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyTArGSerGlnAlaVa 2332
7062 GATCTTCAATATGCGCCCATCACCGGAACAGGCTTTATGACTATAGAAGCCAGCGGT 7121
2332 lGlnGluHisAlaSerThrAsnMetGIyLeuGluAlaIleIleArGlyAlaIleuMetGI 2352
7122 GCAGGAACATGCGACGACCAACATGGGGCTGAGGCCATTAATTAAGAAAGCACTCATGG 7181
2352 yLySPTyAspGlnTrGluGluGlySerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
7182 TAAATATGACCAAGTGGAAAGATCCCGCGCTGACGCGCAATGCTTTTAACCTTGAA 7241
2372 nAlaSerAlaSerLeuProAlaAlaIaMetProIleThrAlaAlaAspGlyArGSerAspHis 2392
7242 TGCAGTGCAGGCTGCGCGCTGTATGCCAATTAACCGCTGAGCAGGAGTGACCA 7301
2392 sThrLeuThrSerProGlyGlyGlyGlyValLyAspValSerGlyArGProSerSerAr 2412
7302 CACACTCACCTCGCCAGGTGGCGGCGGAAAGGCCAAGGTCTTGGGACAGACCCAGCAGCCG 7361
2412 gLyAsAlaLySerProAlaProGlyLeuAlaSerGIyAspArGProProSerValSerSe 2432
7362 AAAAGCCAAGTCCCGGCGCCCGGCTGGATGTGGGAGACGGCACCTCTGTCTCTC 7421
2432 rValHisSerGIuGlyAspCysAsnArGArGThrProLeuThrAsnArGValITrGluAs 2452
7422 AGTGCACTGGAGGAGATGCAACCGCGGACGGCTCACCAACCGCGTGGAGGA 7481
2452 pArGProSerSerAlaGlySerThrProPheProTyTAsnProLeuIleMetArGLeuG 2472
7482 CAGGCCCTGCTCGCAGGTTCACAGCCATTTCCCTTCAACCCCTTATCATCATGGGTGCA 7541
2472 nAlaGlyValMetAlaSerProProProProGlyLeuPProAlaGlySerGIyProLeuAl 2492
7542 GCGGGGTGTATGCTTCCCAACCCCAACGGGCTCCCGCGGAGACGGGGCCCTCTGC 7601
2492 aGlyProHisHisAlaTrPAspGluGluPProLySPProLeuLeuCysSerGIyTrGluTh 2512
7602 TGGCCCCCAACAGCTTGGAAGAGAGCCAAAGCCACTGTCTGTCTGAGATACGAGAC 7661
2512 rLeuSerAspSerGlu 2517


```
QY 361 GIUGINGIUAenleugIuIySGImeArGInleuAlaValIleProProMeIeutyR 400
Db 1660 GAGCGAGGAAACCGGAAAGACAGATGCGCCAGCTGCGCGTGTATCCCGCCATGCTGTAC 1719
QY 401 AAspIaaSPGInGInaArgIleIySPheIleAenMeIaenGIyLeuMeAlaSPProMet 420
Db 1720 GAGCGTGAACGACGCGCATCAAGTTTCATCAACATGAAGAGGACTTATGCGCCAGCCCATG 1779
QY 421 LysValIyTyLysAspArgInValMeIaenMeCTrPSerGIuGIuIySGIuThrPhe 440
Db 1780 AAGGTGTAAAGAACCGCCAGGTCAATGAACATGTGAAGTGAACAGAGAAAGAACCTTC 1839
QY 441 ArgGIuIySPheMeGIuIhiSProIySaenPheGIyLeuIleAlaSerPheIeugIuArg 460
Db 1840 CGGGGAAGTTTCATCCAGCATCCCAAGAACTTTGGCTGTATCGCATTCATTCCTGAGAGG 1899
QY 461 LysThrValAlaGIuCyValLeuIyTyTyTyTyLeuThrIyLysAsnGIuSnTyLys 480
Db 1900 AAGACAGTGGCTGAATGCTCTCTTATTTACTGACTTGAAGATGAAGAACTATAG 1959
QY 481 SerLeuValArgArgSerTyArgArgArgGIyLysSerGIuGln---GInGIuGIuGIu 499
Db 1960 AGCCTGTGAGACGAGAGCTATGGCGCGCGCAAGAGCCACAGCAACAGCAGCAG 2019
QY 500 GInGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 519
Db 2020 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2079
QY 520 LysAspGIuIySGIuIySGIuIySGIuIySGIuIySGIuIySGIuIySGIuIySGIu 539
Db 2080 AAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
QY 540 AsnAspLySGIuAspLeuLeuLySGIuLySThrAspAspThrSerGIyGIuAspAsnAsp 559
Db 2140 AACGCAAGAGAAAGACTCTCTCAAGGAGAAACAGACAGACACTCAGGGAGAGCAACAG 2199
QY 560 GIuIySGIuIyAlaValAlaSerLySGIyArgLyThrAlaAsnSerGIuGIyArgLyS 579
Db 2200 GAGAAAGAGGCTGTGCTGCCCAAGGCCGCAAAACTGCGCAACAGCAGGAGAGAGCAAA 2259
QY 580 GIyArgIleThrArgSerMeAlaAsnGIuIyAsnSerGIuGIuAlaIleThrProGIn 599
Db 2260 GCGCCCATCACCCGCTCAATGACTTAATGAGGCCAACAGCGAGAGGCCATCACCCCCAG 2319
QY 600 GInSerAlaGIuLeuAlaSerMeGIuLeuAsnGIuSerSerArgTyThrGIuGIuGIu 619
Db 2320 CAGACGCGCGAGCTGCGCTCCATGAGCTGAATGAGATTCGCTGAGCAAGAGAGAA 2379
QY 620 MetGIuThrAlaIyLySGIyLeuGIuIySGIyArgLyThrAlaIleAlaArg 639
Db 2380 ATGGAAACAGCCCAAGAAAGGTCTCCTGGAAACAGCGCCGCACTGGCTCGGCCCTCGG 2439
QY 640 MetValGIySerLyThrValSerGIuCyLySAsnPheTyPheAsnTyTyIySArg 659
Db 2440 ATGTGTGGCTCCAAAGCTGTGTGCGAGTGAAGACTTTACTTCACTACAGAGAGAG 2499
QY 660 GInaenLeuAspGIuIleuGIuGIuIhiSlySlySLeuLyMeGIuIySGIuArgAsnAla 679
Db 2500 CAGAACTCTGATGAGATCTTGACAGACGACAACTGAGATGAGAGAGAGAGAGAGAGAG 2559
QY 680 ArgArgLyLyLyLyLyAlaProAlaAlaAla--SerGIuGIuAlaAlaPheProProVa 699
Db 2560 CGGAGAGAAAGAAAGAAAGCGCGCGCGCGCGCGCGAGAGAGAGCTTCCCGCCCT 2619
QY 699 IValGIuSPGInGIuMeGIuAlaSerGIyValSerGIyAsnGIuGIuGIuMeValGI 719
Db 2620 GGTGAGATGAGAGATGAGAGCGCTCGGCGCTGACCGGAAATGAGAGAGAGATGGTGA 2679
QY 719 uGIuIyAGIuAlaIeuhIaIaSerGIyAsnGIuValProArgGIyGIuCySerGIyPr 739
Db 2680 GGAGGCTGAA----- 2689
QY 739 oAlaThrValaenAsnSerSerAspThrGIuSerIleProSerProIhiThGIuAlaAl 759

Db 2690 -GCCACTGTCAACACACTGAGACACAGAGATCCCTCTCTCTCACTGAGGCGCG 2748
QY 759 aLyAspThrGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 779
Db 2749 CAAGAGACACAGGACGAGATGGGCCCAAGCCCGCAGCCACTCGGCGCGAGCGGCGCAC 2808
QY 779 rProGIyProProThrProPro--ArgArgThrSerArgAlaProIleGIuProThrPro 798
Db 2809 CCCCAGGCGCCACCCACCCACCGAGAGCATCCCGGCC--CCCACTGAGAGCCACCGCG 2867
QY 799 AlaSerGIuAlaThGIyAlaProThrProProProAlaProProSerProSerAlaPro 818
Db 2868 GCCTTGAAGCCACCGAGAGCCCTTACGCCCAACAGACCCCACTGCGCTTCACT 2927
QY 819 ProProValAlaProIySGIuGIuIySGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 838
Db 2928 CCTCTGTGTGTCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2987
QY 839 GIuGIyGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 858
Db 2988 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3047
QY 859 GIuGIuProValIySerGIuCySerGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 878
Db 3048 GAGAGCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3107
QY 879 AAspAlaGIuAlaAlaGIuAlaThrAlaGIuGIuAlaIeuhIySAlaGIuIySGIuGIy 898
Db 3108 GAGCGGAGAGCGCGCTGAGGCCAGCGCGAGGGGGGCTCAAGGACAGAGAGAGAGAGAG 3167
QY 899 GIySerGIyArgAlaThrThrAlaIySerSerGIyAlaProGInaSPSerAspSer 918
Db 3168 GAGAGCGCGAGGCGCCACAGCCAAAGAGCTCGGGCGCGCCCAAGAGACAGCACTCCAGT 3227
QY 919 AlaThrCySerSerAlaAspGIuValaAspGIuAlaGIuGIuGIyIyAspLySAsnArgLeu 938
Db 3228 GCTACTGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3287
QY 939 SerProArgProSerLeuLeuThrProThrGIyAspProArgAlaAsnAlaSerProGIn 958
Db 3288 TCCCCAAGGCCAGCTCTCTCAACCCGACTGGCGAGCCCGGGGCAATGCTCACCCAG 3347
QY 959 LysProLeuAspLeuLyGIuLeuLySGIuArgAlaAlaAlaIleProProIleGIuAl 978
Db 3348 AAGCCACTGAGACTTGAACAGCTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3404
QY 979 ThrIySValIhiSGIuProProArgGIuAspAlaAlaIaProThrLySProAlaProPro--Al 998
Db 3405 ACCAAAGTCCATGAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3464
QY 998 aProProProProGInaenLeuGIuProGIuSerAspAlaProGInaProGIySerSe 1018
Db 3465 CCGACCGGCACCGCAAAACCTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524
QY 1018 rProArgGIyLySerArgSerProAlaProProAlaAspLySGIu----- 1033
Db 3525 CCCCCGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3584
QY 1034 -----AlaPheAlaAlaGIuAlaGIuIyLeuProGIyAspProProProProTyTrTh 1050
Db 3585 GTTCTTCCAGCTTCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3644
QY 1050 rSerGIyLeuProPheProValProProArgGIuValIleIySAlaSerProIhiAlaBr 1070
Db 3645 TTCCGGGCTGCGCTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3704
QY 1070 aAspProSerAlaPheSerTyAlaProProGIyIhiAProLeuProLeuGIyLeuIhiSAs 1090
Db 3705 GAGCCCTCAGAGCTTCTCTAGCTGCTCACTGAGTCAACCACTGAGAGAGAGAGAGAGAGAG 3764
QY 1090 pThrAlaArgProValIleuProArgProProThrIleSerAsnProProProLeuIleSe 1110
```

Db 3765 CACTGCCGGCCCGTCTGCGCGCCACCACCATCTCCAAACCCGCTCCCTCATCTC 3824
Qy 1110 rSerAlaIyehiAProSerValIeuGIuaRGlnIleGIyAlaIleSerGlnIyMetSe 1130
Db 3825 CTCTGCCAACACCCCGCGCTCTCGAGAGGCAATAGTGCCATCTCTCCAGGAATGTC 3884
Qy 1130 rValGlnLeuHiIevaIProTyrSerGIuHiBaIAlaIyAlaProValGIyProValIThMe 1150
Db 3885 GGTCCAGCTCCACGTCCTGACTCAGACATGCCAAGGCCCGGTGGGCTGTCAACAT 3944
Qy 1150 rGIyLeuProLeuProMetAerProIyIyIleuAlaProPheserGIyAlIyGlnGI 1170
Db 3945 GGGGCTGCGCTCCCTCCATGGACCCCAAAAGCTGGCACCTTCGCGAGATGAAGCGGA 4004
Qy 1170 uGIuLeuSerProArGIyGlnIaGIyProProGIuSerIeuGIyValProThraIaGI 1190
Db 4005 GCAGCTGTCCCAAGGGGCCAGGCTGGGCCAACCGAAGCTTGGGGTGTCCACAGCCCA 4064
Qy 1190 nGIuAlaSerValIeuArGIyThraIaIeuGIySerValProGIyGIySerIleThry 1210
Db 4065 GAGAGCGTCCGTGCTGAGAGGACAGCTTGCGCTCATTCGGGGCGGAAGCATCACAA 4124
Qy 1210 rGIyIleProSerThraIyAlaProSerAerProAlaIleThThraIyArGIySerIleTh 1230
Db 4125 AGGCATTCCAGACACGGGTGCTCTCGACAGGCGCATCACATACCGCGCTTCATCAC 4184
Qy 1230 rHiSGIyThraIaAPrValIeuTyrIySGIyThraIleThraIleThraIleGIyGluAs 1250
Db 4185 CCAGCGACCCCGAGCTGACCTCTGACAAAGGACCATCACAGGATCATCGGCGAGA 4244
Qy 1250 pSerProSerArGIyLeuAerArGIyArGIyAerSerIeuProIyGIyHiIeValIleTy 1270
Db 4245 CAAGCGAGTCCGTGGACCGCGCGGAGACAGCTGCCAACGGGCCACGTCATCTA 4304
Qy 1270 rGIyGIyIyIyGIyHiIeIeuSerTyrGIyGIyGIyMeSerValIThrGlnCySe 1290
Db 4305 CGAAGCAAGAAAGGCGCATCTTGTCTTAATGAGGGTGCATGTCTGTGACCCAGTGCTC 4364
Qy 1290 rIySGIyAerGIyArGIySerSerGIyProProHiAGIuThraIaIaProIyAerTh 1310
Db 4365 CAAGAGACCGAGAAAGCTCAGGACCCCGCATGAGACGCGCGCCCAAGCGCAC 4424
Qy 1310 rTyrAerMetMetGIuGIyArGIyAlaIyArGIyAlaIeSerSerAlaSerIleGIyGlyLe 1330
Db 4425 CTATGACATGATGAGAGGCGCGGTGGAGAGCATCTCTTCAGCCAGCATCAAGGTCT 4484
Qy 1330 uMetGIyArGIyAlaIleProProGIuArGIyIeSerProHiIeIeuIySGIuGlnHiIe 1350
Db 4485 CATGGGCGGTGCATCCCGCGAGGACACAGCCCCACACCTCAAGAGACGACCA 4544
Qy 1350 rIleAerGIySerIleThGIuGIyIleProArGIySerTyrValGIuAGIuGIuAerPy 1370
Db 4545 CATCCCGGGTCCATACACAAAGGATCCCTCGTCTCATGTAGAGACAAAGAGACTA 4604
Qy 1370 rIleuArGIyGIuAlaIyIeIeuIeuIyArGIyGIyThraProProProProProSe 1390
Db 4605 CCGGCTCGGAGGCAAGCTCTTAAAGGGAGGCAAGCTCCGCGCCCAAGCGCTC 4664
Qy 1390 rArGIyAerLeuThraIaIyIyIyIyThraIaIeGIyProLeuIyIeIeuIyAerProAl 1410
Db 4665 ACGGGACCTGACCGAGGCTTACAAAGACGAGCGCTGGGCGCTCAAGCTGAAGCGGCG 4724
Qy 1410 rHiSGIyGIyLeuValAlaThraIyIySGIyAlaGIyArGIySerIleHiAGIuIleProAr 1430
Db 4725 CCATAGGGGCTGTGTGGCCAGGTGAAGAGGGGCGCTCCATCATGATGATCCCGG 4784
Qy 1430 gGIuGIuLeuArGIyHiIeThraProGIuIeuProIeuAlaProArGIyProLeuIySGIuIySe 1450
Db 4785 CGAGGAGCTGCGGCAACGCGCGAGCTGCCCTCGGCGCGCGCTCAAGAGAGGCGCTC 4844
Qy 1450 rIleThraIyGIyThraProLeuIyTyrAerThraIyAlaIeThThraIyGIySerIyIy 1470
Db 4845 CATCAGCAGGGGACCCGCTCAAGTACGACACCGGCGGTCCACCATGTGGCTCAAAAA 4904

Qy 1470 rHiAerValArGIySerIeuIleGIySerProGIyArGIyThraPhaProProValHiIeProLe 1490
Db 4905 GCACGACGTACGCTCTCTCATCGGACAGCCCGCGGACCTTCCACCGCTGACCCGCT 4964
Qy 1490 uAerValMetValaAPrAlaIyAlaIeuGIuArGIyAlaIyTyrGIyGIuIySerIeuIySe 1510
Db 4965 GATGTGATGGCGGACCGCGGACCTGGAACCTGCTCTTACAGAGAGAGCTGAAGAG 5024
Qy 1510 rArGIyProGIyThraIaSerSerGIyGIySerIleAlaArGIyAlaProValIleIeva 1530
Db 5025 CCGGCAAGGACCGCACAGCTCGGGGGCTCCATTGCGGCGCGCGCGCTCATTTGT 5084
Qy 1530 lProGIuIeuGIyIyProArGIyInSerProIeuThraIyGIyAerHiAGIyAlaProPh 1550
Db 5085 GCCTGAGCTGGGACCGCGGACAGAGCCCTTACCTATGAGAGACACGGGACACCTT 5144
Qy 1550 eAlaGIyHiIeIeuProArGIySerProValIThMeArGIyIuProThraProArGIeGI 1570
Db 5145 TGCCGCGACCTCCACAGAGTTTCGCCGTGACCAAGCGGAGGCCACGCGCGCTGCA 5204
Qy 1570 nGIuGIySerIeuSerSerIyAlaSerGIuAerArGIyIeIeuThraIySerThraProAr 1590
Db 5205 GAGAGCGAGCTTTCGTCCAGAAAGCATCCAGAGCCGAAGCTGACGTGACGCTCG 5264
Qy 1590 gGIuIleAlaIySerProHiIeSerThraValProGIuHiIeHiIeProHiIeSerPr 1610
Db 5265 TGAATGCGCAAGTCCCGGACAGCACGTCGCGGAGCACACACACCCCATCTTGCC 5324
Qy 1610 oTyrGIuHiIeIeuIeuArGIyAlaIeGIyAlaIeArGIyAlaIeArGIyArGIySerHiIeProLe 1630
Db 5325 CTATGAGACACTGCTTCGGGCGTGAAGTGGCGTGAACCTGTATCGAGCCACATCCCT 5384
Qy 1630 uAlaIeAerProThraIeIeProArGIyIleProIeuAerAlaIaIaIaIaIyTyr 1650
Db 5385 GGCCTTGACCCGACCTCATACCCGCGGACATCCCTGAGAGCAGCGCTCCACTA 5444
Qy 1650 rIleuProArGIyIeIeuAlaIeProAerProThraIyProHiIeIeuTyrProProTyrIleuI 1670
Db 5445 CTGCGCCGACACTGGCCCCCAACCCCACTACCCGCACTTACCCACTTACTCAT 5504
Qy 1670 eArGIyTyrProAerThraIaIaIeGIuAerAerGIuThraIeIeAerAerPyI 1690
Db 5505 CCGGCGCTACCCGACAGCGCGCGCTGAGAAACGGCAGACCATCATCAATCATCAT 5564
Qy 1690 eThraSerGIuIyMeHiIeAerThraIaIeAlaIeGIuAerAlaIaIaIaIaIe 1710
Db 5565 CACTTGACAGATGACACACACCGGCGCACCGGCATGCGCCAGCGAGCTGATGTCT 5624
Qy 1710 uArGIyIeIeuSerProArGIySerSerIeuAlaIeAerTyrAlaIaGIyProArGI 1730
Db 5625 GAGGGGCTCTGCGCCCGGAGTCTCGCTGGCACTCAACCTAGCGTGGGCTCCGAGG 5684
Qy 1730 yIleIleAerIeuSerGIuValProHiIeIeuProValIeIeuAlaIeProThraProGIyTh 1750
Db 5685 CATCATGACACTGTCCCAAGTGACACCTGCTGTGTGTGCGCCCGACACAGGAC 5744
Qy 1750 rProAlaThraIaIeAerArGIyAlaIyIeIeuProThraIaIeProGIuProPheserSe 1770
Db 5745 CCAAGCACCGGCAATGAGACCGCTTGCCTTACCTCCACCGGCGCCAGCCCTTCAGCAG 5804
Qy 1770 rArGIyIeSerSerProIeuSerProGIyGIyProThraIeIeuThryIyAerProThra 1790
Db 5805 CCGCACAGACGCTCCCACTCTCCCAAGAGGTCCAAACACTTACAAACCAACAC 5864
Qy 1790 rThraSerSerGIuArGIyArGIyAerArGIyAlaIyAerArGIyAerArGIyAer 1810
Db 5865 CAGTCTCTGTCGAGCGGAGAGAGACCGGATGAGAGCGGAGACCGGAGATCGGAGCG 5924
Qy 1810 gGIuIySerIleIeuThraIeIeuThraIyThraIyAlaIeAlaIeProIleTyrArGIProGI 1830
Db 5925 GAAAAAGTCAATCTCATGTCACCAAGAGGTGAGACGAGCCATCTGAGAGACTGG 5984

QY 1830 YThrGluInserSerGlySerSerGlyGlyGlyGlySerSerSerAr 1850
DB 5985 TACAGAGCAGAGCAGCGGAGAGCGGAGCGCGGGGTGGGGGAGCAGCAGCGCG 6044
QY 1850 GProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
DB 6045 CCCCCCTCCCACTCCCACTGCCACAGACTGCCCACTCCCTCCGAGCCAGAGATGC 6104
QY 1870 AleuGlnGlnArgProSerValIleuHisAenThrGlyMetLysGlyLeIleThrAlaVa 1890
DB 6105 CCTCCAGAGAGAGACCAAGTGTCTTCAACAACAAGGCATGAAGGATATATACACGCTGT 6164
QY 1890 GLeuProSerLysProThrValIleu-----ArgSerThrSerSerSerProVa 1907
DB 6165 GGAGCCACGACGACGCCACGGTCTGTAGGTCGAGGTCACCTCCACTCTCCACCGCT 6224
QY 1907 LArgProAlaAlaThrPheProProAlaThrHisGlyProLeuGlyGlyThrLeuAspG 1927
DB 6225 TCGCCCGCTGCCAATCCCACTGCCACCACTGCCACCTGGGGGCGACCTCGATGG 6284
QY 1927 ValTyProThrLeuMetGluProValIleuLeuProLysGluAlaProArgValAlaAr 1947
DB 6285 GGTCACCTCACTCACTATGAGACCGCTCTTGTCTGCCAAGAGGCCCGGGGTGGCGCG 6344
QY 1947 GProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
DB 6345 GCCAAGAGCGGCGCCGAGCAGACCGGCGCATGTCTCTCCGCAAGCCCGCCGCGCTC 6404
QY 1967 GLeuGluGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProPr 1987
DB 6405 CCGGCTGGAGCGCGCTCTCTCCCAAGAGGAGCTGGAGCGCGCGCTCTGTGCTCTC 6464
QY 1987 GValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAl 2007
DB 6465 TGTCTCTGGCCAGCGCACCACATGCCCGCACCTCTGGAGAAACCTGGACCTCACACGCC 6524
QY 2007 AserProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGl 2027
DB 6525 CAGCCCGGAGCGCGCGCGCGCACTGGCTCGGCTCGGACCGCGCAAGGAAAGACTCA 6584
QY 2027 nSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
DB 6585 AAGTAAACCTTTTCATCTCAGAACTGGAACTCGCTTCTGTGGTTACACAGCGCAG 6644
QY 2047 rTySerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAs 2067
DB 6645 CTACAGCCCCGAAAGGGGTGAGCCCGCTCAGCCCTGTGAGCTCACCAAGCTACACCA 6704
QY 2067 PlysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyLysLeuAr 2087
DB 6705 CAAGGGGCTCCCCAAGCACCCTGGAAGCTCGCAAGAACCACTGGAAGGGGAGCTGCG 6764
QY 2087 GProLysGlnProGlyProValLysLeuGlyGlyValAlaHisLeuProHisLeuAr 2107
DB 6765 GCCCAAGCAGCAGCGCGCTTGGCAGGAGGCGCGCCCACTCCCAACCTGGCG 6824
QY 2107 GProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
DB 6825 GCCGCTGCTGAGAGCCAGCCCTGTCTCAGCCCGCTGTCTCCAGACCGCCCGGGGTCAA 6884
QY 2127 sGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTy 2147
DB 6885 AGGTACACGAGGGGTGTCACTGGCCAGCAGCATGAGAGTATACACAGAGACTA 6944
QY 2147 rThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerTherAr 2167
DB 6945 CACCCGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7004
QY 2167 OGlyAlaSerCysProValIleuAspLeuArgArgProProSerAspLysLeuTyrLeuProPr 2187
DB 7005 TGGGGCAGACTGCCCTGTCTGAGACTTCGGCGGCCCAACCAAGTACCTTACCTCCGCGC 7064
QY 2187 oProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerPr 2207

DB 7065 CCGGAGCATGATGCCCCCGCGCTCCCGCCCAAGCGGAGGCGCAAGGATCTCC 7124
QY 2207 OGluProAsnLysThrSerValIleuGlyGlyGlyLysAspGlyIleGluProValSerPr 2227
DB 7125 AAGCCCAACAGACGCTGGCTTGGGTGTGTGAGAGCGGATATTGAACGCTGTCCCG 7184
QY 2227 oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrAr 2247
DB 7185 ACCGAGGGCATGACGAGCAGCGGCACTCCCGGAGTGTGTATCCGCTCTGTACCG 7244
QY 2247 GAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGl 2267
DB 7245 GGATGGGAGACGAGCGAGCCAGAGATGGGCTTCAAGTCTCCAGCGCAACACAGCCA 7304
QY 2267 nProProAlaPhePheSerLysLeuThrGlySerAsnSerAlaMetValLysSerLysLy 2287
DB 7305 GCCGCAAGCTTCTTACAGCACTTACGAGAGCACTCCGCAATGTCTCAAGTCCAGCA 7364
QY 2287 sGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSe 2307
DB 7365 GCAAGAGATCAACAAGAAAGCTGAACACCAACCGGAATGAGCTGAATACAAATATCAG 7424
QY 2307 GlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTy 2327
DB 7425 CCAGCTGGAGCGGAGATCTTCAATATGCGCGCATCACCGGAACAGGCTTATGACCTA 7484
QY 2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleAr 2347
DB 7485 TGAAGCCAGCGGTGTGAGAGAACATGCCAGCAACAACTGGGGCTGAGAGCCATATTAG 7544
QY 2347 GlysAlaLeuMetGlyLysTyrAspGlnTTPGluGluSerProProLeuSerAlaAsnAl 2367
DB 7545 AAAGGCACTCATGGGTAAATATGACCAAGTGGAAAGTCCCGCGCTCAGCGCCAATGC 7604
QY 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlAs 2387
DB 7605 TTTTAACTCTGTAATGCGCAGGCGCGCTGCTGTATGCCATTAACGCTGTGA 7664
QY 2387 PGLyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGl 2407
DB 7665 CGGACGAGTGAACAACACTACCTCCGAGGTGGCGGGGAAAGCCAAAGTCTCTGG 7724
QY 2407 YArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgPr 2427
DB 7725 CAGACCCAGACGCGCAAAAGCTCCCGGCGCTCGGATCTGGGAGACCGGCGC 7784
QY 2427 oProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrs 2447
DB 7785 ACCCTGTCTCTCCTCAGTGCATCGAGGAGAACTGCAACGCGGAGCCGCTCACCA 7844
QY 2447 nArgValTTPGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLe 2467
DB 7845 CCGCTGTGGAGAGCAGGCGCTCGTCCGAGGTTCCAGCCATTCCTTACACCCCT 7904
QY 2467 uIleMetArgLeuGlnAlaGlyValIleMetAlaSerProProProProGlyLeuProAlaGl 2487
DB 7905 GATCATGGGGCTGCAAGCGGGGTGTATGCTTCCCAACCCCAACCGGGGCTCCCGCGGG 7964
QY 2487 ySerGlyProLeuAlaGlyProHisHisAlaThrAspArgGluGluProLysProLeuLeuTy 2507
DB 7965 CAGCGGGCGCTCGCTGGCGCCCAACAGCTGGGAGCGAGGAGCCCAACCACTGCTGTG 8024
QY 2507 sSerGlnTyrGluThrLeuSerAspSerGlu 2517
DB 8025 CTGCACTGACGACACTCTCCGACAGCGAG 8055

RESULT 15

ADQ23294/c
ADQ23294 standard; DNA; 9079 BP.ADQ23294;
AC
XX

DT 26-AUG-2004 (first entry)
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6114.
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM db.
XX Homo sapiens.
OS
PN MO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
DR
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 6114; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 9079 BP; 1268 A; 2703 C; 3184 G; 1924 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 9079
Score: 12845.50 Matches: 2473
Percent Similarity: 98.10% Conservative: 6
Best Local Similarity: 97.86% Mismatches: 20
Query Match: 97.20% Indels: 31
DB: 12 Gaps: 3

US-09-522-753-5 (1-2517) x ADQ23294 (1-9079)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 8552 ATGTCCGGGCTCCACACAGCTGTGGCAGACGTTGAGGGCCACTGAGCCCCGTACCCG 8493
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 8492 CCCACAGCCTTTCCTAACCAAGTCAGATGCGCCGAGCCGACACGACGTCGGGCTCTCG 8433
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 8432 GAGTACACAGACACACTCCGCGACTATGCTCCACCTGTGCGCCGCTCCATCATCCAG 8373
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyLysGlnArgSerGln 80
Db 8372 CCCACGCGGAGGCGCTCTCCCTGCTGTCTGAGTTCCAGCCCGGAGATGACGTTCCAG 8313
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100

Db 8312 GAGCTCCACCTGCGGCGCAGAGATCCCACTCATACCTGCGGAGCTGGGAACTGAGAGATG 8253
QY 101 GluPheIleGluSerTyrArgProArgLeuGluLeuProArgProLeuLeuArgPro 120
Db 8252 GAGTTCATTGAAACCAAGCCGCTCGCTAGAGCTGCTGCTGACCCCTCTCTGGACCG 8193
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 8192 TCACCCCTGCTGGCCAGCGGCGACCTCGCGGACTGTGAAGACTTCACCAAGACCGGCC 8134
QY 141 LeuThrGlyLysLeuGluProValSer-ProProSerProProHisThrAspProGluLe 160
Db 8133 CTGACGGGCAAGCTGGAACCGGTGTCTCCACCCAGCCGCCGACACTGACCTGTGAGCT 8074
QY 160 UGULeuValProProArgLeuSerTyrGluGluLeuIleGlnMetMetAspArgValAs 180
Db 8073 GGAAGCTGTGCGCGCAGCGCTGTCCAGAGGAGACTATCCAGAACATGACCCGCTGGA 8014
QY 180 PArgGluIleThrMetValGluGlnGlnIleSerTyrLeuLysLysGlnGlnIle 200
Db 8013 CCGAGAGATCACCATGTGTAGCGACGATCTTAAGCTGAGAGAAAGACGACACAGCT 7954
QY 200 UGUUGUUGUAlaAlaLysProProGluProGluLysProValSerProProIleG 220
Db 7953 GAGAGAGAGAGCTGCGCAAGCCGCGGAGCTGAGAAACCGGTGTACCGCGCCCATCGA 7894
QY 220 USerTyrHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValAlaGlu 240
Db 7893 GTCGAACACCGCAGCTGTGTGAGATCATTCAGAGAGAAACCGAAGAGCTGTAGAC 7834
QY 240 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPr 260
Db 7833 TGCATATCGATTTCTGAAGGCTTGAGGCCCTCCAGGTGAGTGCCTGTGACCAAGACC 7774
QY 260 OSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLe 280
Db 7773 CTCGACACCCGGAGATCATATGAAATCAAAATTAACACGAGCGATGCGAAGAAAGCT 7714
QY 280 UileLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCyG 300
Db 7713 AATCTGTACTTCAAGAGGAGAAATACGCTCGGAAACAATGGAGACAGAAATTTCTCCA 7654
QY 300 nArgTyrAspGlnLeuMetGlnAlaLeuGluLysLysValAlaGluArgIleGluAsnAsnPr 320
Db 7653 GCGCTATGACCACTCATGAGGCTGTGGAGAAAGAGTGAAGCATTCAGAAACAACC 7594
QY 320 OATGATGATGAlaLysGluSerLysValAlaArgIuTyrGluLysGlnPheProGluI 340
Db 7593 CCGGCGGCGGCGCAAGAGAGCAAGGTTGCGAGTACTTACGAGAACAGATTCCTGAGAT 7534
QY 340 eArgLysGlnArgGluLeuGlnGlnLysMetGlnSerArgValGlyGlnArgGlySerG 360
Db 7533 CCGCAAGACGCGGAGCTGTGAGAGGCGCATGACAG--AGGATGGCGCAGCGGCGCACTGG 7477
QY 360 YLeuSerMetSerAlaAlaArgSerGlnHisGluValSerGluIleIleAspGlyLeuSe 380
Db 7476 GCTGTCTCATGTCCCGCCCGCCGACGACGACGAGGTGTCAAGATTCATCATGCTCTC 7417
QY 380 TGIUGInGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTy 400
Db 7416 AGAGCAGAGAACTGTGAAGAACAGATGCGCAGCTGCGCTGATCCGCCCATGCTGTA 7357
QY 400 rAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
Db 7356 CGACGCTGACCAACACGACATCAGTTTCATCAACATGAAGAGGCTTATGCGACCCCAT 7297
QY 420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db 7296 GAAAGGTATCAAAAGACCGCAGGTCTATGAACATGTGAGTGAAGCAGAGAGAGACCTT 7237
QY 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleIleSerPheLeuGluAr 460

Db 7236 CCGGAGAAAGTTTCATGCAGCATCCCAAGAACTTTGGCTGATCGCATCTTCCTGGAGAG 7177
Qy 460 glythrValaIaGIuCyseValLeuTyrTyrTleuThrylyseAaspGluAaspTyrly 480
Db 7176 GAAGACAGTGGCTGAGTGGCTCTTATTACCTGACTGAAGAAAGAACTATATA 7117
Qy 480 sSerleuValaIaGAsSerTyrArGArGArGlylyseSerGlnGlnGlnGlnGlnGlnGln 500
Db 7116 GAGCTGGGTGAGACGGAGCTATCGGGCGCGGCAAGAAC--CAGCACAAACAAAG 7059
Qy 500 ngInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 7058 CCGGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 6999
Qy 520 sAspGluysGluysGluysGluysGluysGluysGluysGluysGluysGluysGluys 540
Db 6998 AATATGAAAGAGAAAGAAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6939
Qy 540 nAspLyseGluAaspLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 560
Db 6938 CGACAAAGAAAGCTCTCTCAAGAGAAAGACAGACAGACCTCAAGGAGAGACAGACAG 6879
Qy 560 uLyseGluValaIaIaSerLyseGlyYArGlyThrAlaAaspSerGlnGlyArGlyseG 580
Db 6878 GAAGGAGGCTGTGGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6819
Qy 580 YArGlyeThrArGserMetAlaaspGluAaaspSerGluGluAaIleThrProGlnG 600
Db 6818 CCGCATCACCGCTCATATGATGAGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAG 6759
Qy 600 nSerAlaGluLeuAlaSerMetGluLeuAaspGlnGlnGlnGlnGlnGlnGlnGlnGln 620
Db 6758 GAGCGCGCAGCTGGCTCTCATGAGAGTGAATGAGATTCCTCGTGAAGAAAGAAAT 6699
Qy 620 rGluThrAlaLyseGlyLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 640
Db 6698 GGAACAGGCAAGAAAGGTCTCTCTGGAACACGCGCCCAACTGGTCGCGCATCCCG 6639
Qy 640 tValGlySerLyseThrValSerGlnCyseLyseAaspPheTyrPheAaspTyrLyseYArG 660
Db 6638 GGTGGGCTCAAGACTGTGTGCGAGTGTAAAGACTTACTTAACTTACAAAGAGAGGA 6579
Qy 660 nAaspLeuAaspGluLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 680
Db 6578 GAACTCGATGAGATCTTGAGAGACAGACAGCTGAAGATGAGAGAGAGAGAGAGAG 6519
Qy 680 gArGlyLyseLyseLyseAlaProAlaAlaIaSerGluGluAaIaPheProProValYa 700
Db 6518 GAGGAAGAGAGAAAGAGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 6459
Qy 700 lGluAaspGluGlnMetGluAaSerGlyValaIaSerGlyAaspGluGlnGlnMetValG 720
Db 6458 GAGAGATGAGAGATGAGAGCGTGGGCGTGAAGGAGAAATGAGAGAGAGATGGTGA 6399
Qy 720 uAlaGluAlaLeuHiAlaSerGlyAaspGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 740
Db 6398 GCGTGA-----GC 6390
Qy 740 aThrValAaspAaspSerAaspThrGlnSerIleProSerProHisThrGluAlaAlaLy 760
Db 6389 CACTGTCAACAAACGCTCAGACACCAAGAGACATCCCTCTCTCACACTGAAGCGCCCA 6330
Qy 760 sAspThrGlyGlnAaspGlyProLyProProAlaAthrLeuGlyAlaAaspGlyProPro 780
Db 6329 GAGACAGGCGCAAAATGGGCCCAAGCCCAAGCCCACTTGAGGCGAGCGAGCGAGCC 6270
Qy 780 oGlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAla 800
Db 6269 AGGGCCACCCACCCACACCGAGAGACATCCGGCC-CCCACTGAGTCACCCCGGCT 6211
Qy 800 eTcGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro 820
Db 6210 CTGAAGCCACTTAGCGCCCTACGCCCCCAAGACACCCCACTTCCCTCTTCACTCTC 6151

Qy 820 roValValProLyseGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 840
Db 6150 CTGTGCTCCCAAG 6091
Qy 840 lYglGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 860
Db 6090 GGGAG 6031
Qy 860 lUpProValLyseSerGluCyseThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 880
Db 6030 AGCCCGTCAAG 5971
Qy 880 lAGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLyseAlaGluLyseGlnGlyYS 900
Db 5970 CCGAGGCGCTGAGGCGACCGCGCAGAGAGGCGCTCAAGCAGAGAGAGAGAGAGAG 5911
Qy 900 eTcGlyArGAlaThrThrAlaLyseSerGlyAlaProGlnAaspSerAaspSerAlaT 920
Db 5910 CCGGAGAGGCGCACACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5851
Qy 920 hTcYseSerAlaAaspGluValaAaspGluAlaGluGlyGlyAaspLyseAaspGlyLeuSerP 940
Db 5850 CCTGAGATGACAGAGAGTGAATGAAGCCGAGAGCGGCGACAGAAACCGGCTGCTCC 5791
Qy 940 rArGProSerLeuLeuThrProThrGlyAaspProArGAlaAaspAlaSerProGlnLyseP 960
Db 5790 CAAGGCGCAGCTCTCTCAACCGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 5731
Qy 960 rLeuAaspLeuLyseGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 980
Db 5730 CACTGGACTTGAAGCAGCTGAAGCAGCAGCGAGCGGTCCATCCCCCATTCAGGTACA 5671
Qy 980 yValHiAGluProProArGAlaAaspAlaAlaProThrLyseProAlaProProAlaPro 1000
Db 5670 AAGTCCATAGAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5611
Qy 1000 rOProProGlnAaspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1020
Db 5610 CCGACCGCAAAACCTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5551
Qy 1020 rGlyLyseSerArGserProAlaProProAlaAaspLyseGlu----- 1033
Db 5550 GGGGCAAG 5491
Qy 1034 -----AlaPheAlaAlaGluAlaGlnLyseProGlyAaspProProCyseThrThrSerG 1052
Db 5490 TCCAGCTTCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5431
Qy 1052 lYleuProPheProValProProArGAlaValIleLyseAlaSerProAlaAlaProAasp 1072
Db 5430 GCTGCGCTTCCCGCGAG 5371
Qy 1072 rOserAlaPheSerTyrAlaProProGlyHiAProLeuProLeuGlyLeuHiAaspThra 1092
Db 5370 CCTCAGCCTTCTCTCTGAGCTTCACTGATCAGCAGCAGCAGCAGCAGCAGCAGCAG 5311
Qy 1092 lAarProValLeuProArGProProThrIleSerAaspProProLeuAlaSerSerA 1112
Db 5310 CCGGCGCGCTTCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5251
Qy 1112 lAlaYhiAaspSerValLeuGlnArGlnIleGlyAlaIleSerGlnGlyMetSerValG 1132
Db 5250 CCAAGACCCCGAGCGTCTCGAGAGCAATATAGTGCATCTCCAAAGAAATGCGGTCC 5191
Qy 1132 lInLeuHiValProTyrSerGlnHiAlaAlaAlaProValGlyProValThrMetGlyL 1152
Db 5190 AGCTTCACGTCCTGTAAGTCAAGAGATGCCAAGAGAGAGAGAGAGAGAGAGAGAG 5131
Qy 1152 eUpProLeuProMetAaspProLyseLyseLeuAlaProPheSerGlyValLyseGlnGlnL 1172
Db 5130 TGCCTCTGCGCAAG 5071

QY 1172 euserProArGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlu 1192
DB 5070 TGTCGCCACGCGGCGCAGGCTGTGGCCACCGGAGAGCTGTGGGTGCCACAGCCAGGAGG 5011
QY 1192 laSerValLeuAArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIleGly 1212
DB 5010 GGTCCGTGCTGAGAGGAGCAGCTCTGAGCTCAgTTCGGGGGGAAGATCCAAAGCA 4951
QY 1212 leProSerThrArGValProSerAAsPeralIleThrTyArGlySerIleThrIleG 1232
DB 4950 TTCCACACACACGCGGTCCCTCCGACAGCCCATCATACACGCGGTCCATCCACAG 4891
QY 1232 lyThrProAlaAsPValLeuTyIleGlyGlyThrIleThrArgIleIleGlyLeuAsP 1252
DB 4890 GCACGCGCAGTGAAGCTCTGTACAAAGGACCATCACAGGATCATCGGCGAGGAGCC 4811
QY 1252 roSerArGLeuAsPArGlyArGlyLeuAsPserLeuProLyGlyValIleTyIleGly 1272
DB 4830 CGAGTCGCTTGACCGCGCGCGGAGAGCAGCTGCCCAAGGCGCAGTCACTACAGAG 4771
QY 1272 lyIleValSGlyHisValLeuSerTyIleGlyGlyMetSerValThrGlnCysSerTy 1292
DB 4770 GCAGAGAGGCGCAGCTTGTCTGTATAGAGGTGACATGTCTGTACCAAGTCTCAAG 4711
QY 1292 luAsPGLyArGSerSerSerGlyProProHisGluThrAlaAlaProLyAsPThrTy 1312
DB 4710 AGGACGCGCAGAGAGCTCAGAGCCCCCATGAGAGCGCGCGCCCAAGGAGCCTATG 4651
QY 1312 sPmetMetGlyGlyArGValGlyArGAlaIleSerSerAlaSerIleGlyGlyLeuMet 1332
DB 4650 ACATGAGAGGAGCGCGGTGGGAGAGCATCTCTCAGCAGCATGAGAGGTCTCATGG 4591
QY 1332 lyArGAlaIleProProGlyUAArgHisSerProHisIleuLyGlyGlnHisIle 1352
DB 4590 GCGGTGCATCCGCGCGGAGCAGACGCCCACTCAAGAGCAGAGCAGCAGCAGCAGC 4531
QY 1352 rgGlySerIleThrGlnGlyIleProArGSerTyValGlyAlaGlnGlyUAAsPTr 1372
DB 4530 GCGGTCTCATCACACAGAGGATCCCTGCTCATCGTGAAGGACAGAGAGCTACCTG 4471
QY 1372 rGArGlyUAAlaLyLeuLeuLyArGlyGlyIleThrProProProProProSerAr 1392
DB 4470 GTGCGGAGGCGCAAGCTCTAAAGGAGGAGGAGCGCTCCGCCCCACCGCTCACG 4411
QY 1392 sPLeuThrGlyUAAlaTyIleThrGlnAlaLeuGlyProLeuLyLeuLyProAlaHis 1412
DB 4410 ACCTGACCGAGGCGCTACAGACGAGGCGCTGGCGCTCCAGACTAAGCGGCGCATG 4351
QY 1412 luGlyLeuValAlaThrValLyGlyUAAlaGlyArGSerIleHisGlyIleProAr 1432
DB 4350 AGGCGCTGTGGCCACGCTGAGAGAGGCGGCGCTCATCATGAGATCCCGCGGAGG 4291
QY 1432 luLeuArGHisThrProGlyLeuLeuProLeuAlaProArGProLeuLyGlySerIle 1452
DB 4290 AGCTGGGCGCAGCGCGGAGCTGCCCTGGCGCGCGCTCAAGAGAGGCTCCATCA 4231
QY 1452 hisGlnGlyThrProLeuLyTyAsPThrGlyAlaSerThrThrGlySerTyIleHis 1472
DB 4230 CGAGAGGACCGCGCTCAAGTACGACCGCGGCTCCACCATGCTCCAAAGAGCAG 4171
QY 1472 sPValArGSerLeuIleGlySerProGlyArGThrPheProProValHisProLeu 1492
DB 4170 ACCTAGCGCTCCCTCATCGGACGCCCGCGGAGCTTCCACCGCTGAGATG 4111
QY 1492 alMetAlaAsPAlaArGAlaLeuGlyUAArgAlaCysTyIleGlyLeuLySerAr 1512
DB 4110 TGTATGCGCGAGCGCGGAGCTGAAAGTGTCTGCTACGAGAGAGCTTAAAGCGG 4051
QY 1512 roGlyThrAlaSerSerSerGlyGlySerIleAlaArGlyAlaProValIleValPro 1532
DB 4050 CAGGAGCGCGCAGAGTGTGGGGGCTCATTCGCGCGCGCGCGCTCATTCGTGCTG 3991
QY 1532 luLeuGlyLyProArGlyInSerProLeuThrTyGluAsPHisGlyAlaProPheAla 1552

DB 3990 AGCTGGGTAGCCGCGCAGAGCGCCCTGACCTATGAGAGCAGCGGAGCAGCTTTCGCG 3911
QY 1552 lyHisLeuProArGlySerProValThrMetArGlyUpProThrProArGLeuGlnG 1572
DB 3930 GCACCTTCCACAGAGTTCCTCCGTGACATGCGGAGGCCACCGCGCTGAGAGG 3871
QY 1572 lySerLeuSerSerSerValAsPArgGlyAlaSerGlnAsPArgLySleuThrSerThr 1592
DB 3870 GCAGCTTGTGTCCAGAGAGCATCCAGAGACGAAAGCTGACCTGACGCTGTGAGA 3811
QY 1592 leAlaLySerProHisSerThrValProGlyHisIleProHisProIleSerProTy 1612
DB 3810 TCCCAAGTCCCGCAGACAGCAGTCCGAGACACCAACCCCATCTCGCCTATG 3751
QY 1612 luHisLeuLeuAArgGlyValSerGlyValAsPLeuTyArGSerHisIleProLeuAla 1632
DB 3750 AGCAGCTGTCCGAGGTGAGTGTGAGCTGTATCGACACCATCCCTGGCT 3691
QY 1632 heAsPProThrSerIleProArGlyIleProLeuAsPAlaAlaAlaTyIleTyLeuP 1652
DB 3690 TCGACCCCATCTCATACCCCGGCGATCCCTTGAGCGAGCGCGCTTACTACTG 3631
QY 1652 roArGHisLeuAlaProAsPProThrTyIleProHisLeuTyIleProProTyIleuIle 1672
DB 3630 CCGGACACTGGCGCCCAACCCCACTACCCGACCTGACACCTTACTATCCGCG 3571
QY 1672 lyTyIleProAsPThrAlaAlaLeuGlyUAAsPArgGlnThrIleIleAsPArTyIle 1692
DB 3570 GCTACCCCGACAGCGCGCGCTGTAGAACCGGAGACATCATCAAGATCACTCACT 3511
QY 1692 eGlnGlnMetHisIleAsPThrAlaThrAlaMetAlaGlnArGAlaAsPMetLeuAr 1712
DB 3510 CGAGCAGATGACACACACACGCGCATGCGCCAGAGAGCTGATGCTGAGG 3451
QY 1712 lyLeuSerProArGlyLeuSerSerLeuAlaLeuAnTyIleAlaGlyProArGlyIle 1732
DB 3450 GCTCTGCGCGCGAGTCTCTGCTGCACTCAACTACGCTGGCGGCTCCGAGGATCA 3391
QY 1732 leAsPLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThr 1752
DB 3390 TCGACCTGTCCAAAGTGCACACCTGTGTGTGTCGCGCCGACACCGAGCAGCC 3331
QY 1752 laThrAlaMetAsPArGLeuAlaTyIleuProThrAlaProGlnProPheSerAr 1772
DB 3330 CACCGCGCATGACCGCGCTTGTCTACTCCCAACCGCGCGCTTACAGAGCGGCC 3271
QY 1772 isSerSerSerProLeuSerProGlyIleProThrHisIleuThrLyProThrThr 1792
DB 3270 ACAGAGAGTCCCACTCTCCCGAGAGTCCAAACACTGAGCAAAACCAACCAACG 3211
QY 1792 eTrSerSerGlyUAArgAlaArGAsPArGlyUAArgAsPArGAsPArGlyUAArg 1812
DB 3210 CTTGTCTCGAGCGGAGGAGCAGACCGGATGTGAAGGAGGACCGGATCGGAGGAAA 3151
QY 1812 ysSerIleLeuThrSerThrThrThrValGlyHisIleArProIleTPArGProGly 1832
DB 3150 AGTCCATCTCATAGTCAACACGAGTGAAGACCGACCCATCTGAGACTGTGACAG 3091
QY 1832 luGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArPro 1852
DB 3090 AGCAGAGAGCGGACAGAGCGGAGCAGCGCGGAGTGGGAGCAGACAGCGCGCCG 3031
QY 1852 laSerHisSerHisIleHisIleHisIleSerProIleSerProArGThrGlnAsP 1872
DB 3030 CTTCCCATCTCCATGCGCACAGACTCGCCATCTCTCCCTCGGAGCCAGAGATCCCT 2971
QY 1872 luGlnArGProSerValLeuHisAsPThrGlyMetLyGlyIleIleThrAlaValGly 1892
DB 2970 AGCAGAGACCGAGTGTGCTTACACACAGAGCATGAAGGTATCATACCGCTGTGAG 2911
QY 1892 roSerLyAsPThrValLeuArGSerThrSerThrSerSerProValArProAlaAla 1912

Db 2910 CCAGAAAGCCAGGCTCTGAGGTCCACTCCACTCTACCCGTTGCGCCAGCTGCCA 2851
Qy 1912 hrPheProRoAlaThrHisCyProLeuGlyGlyThrLeuAspGlyValIYrProThrL 1932
Db 2850 CATTCCCACTGCGCCACCACTGCGCCACTGGCGGCACTCTCGATGGGGTCTACCCCTACC 2791
Qy 1932 euMeGluProValLeuLeuProLYeGluAlaProArgValAlaArgProGluArgPro 1952
Db 2790 TCATGGAGCCCGCTTGTGCTGCCAAGAGAGGCCCGCGGGTCCGCCGAGAGCGGCCCC 2731
Qy 1952 rGAlaAspThrGlyHisAlaPheLeuAlaLYsProProAlaArgSerGlyLeuGluProA 1972
Db 2730 GAGCAAGACACCGGCGCACTGCTTCTCGCAAGCGCCAGCGCTCCGGAGCTGGAGCCCG 2671
Qy 1972 laSerSerProSerLYsGlySerGluProArgProLeuValProProValSerGlyHisA 1992
Db 2670 CCTCTCCCCCGCAAGAGGCTCGAGCCCGCGCCCTTAGTGCTCTGTCTGCGGCAAG 2611
Qy 1992 laThrIleAlaArgThrProAlaLYsAsnLeuAlaProHisHisAlaSerProAspProp 2012
Db 2610 CCACCAATCCCGCGCACCCCTGCGAAGAACCTCGCACTCACACCGCCAGCCCGAGCCCG 2551
Qy 2012 roAlaProProAlaSerAlaSerAspProHisAlaArgIuLYsThrGlnSerLYsProPhS 2032
Db 2550 CGGCGCACCTGCTGCGCTCGAGCCGACCGGAGAAAGACTCAAGATAACCTTTT 2491
Qy 2032 erIleGlnGluLeuGluLeuArgSerLeuGlyTYrHisGlySerSerTYrSerProGluG 2052
Db 2490 CCATCCAGGAACTCGAAGCTCCGTTCTCTGGGTTTACACGGCAGCACTACGCCGGAAG 2431
Qy 2052 lYValGluProValSerProValSerSerProSerLeuThrHisAspLYsGlyLeuProL 2072
Db 2430 GGGTGAAGCCCGTCAGACCTGTGAGCTCACCCAGTGTGACCCACGACAAAGGGCTCCCA 2371
Qy 2072 YHisIleuGluGluLeuAspLYsSerHisIleuGluGlyGluLeuArgProLYsGlnProG 2092
Db 2370 AGAAGCTGGAAGAGCTGCAAGAGCCAGCTGAGGGAGAGCTGCGGCCCAAGACCCAG 2311
Qy 2092 lYProValLYsLeuGlyGlyGluAlaAlaHisIleuProHisIleuAspProLeuProGluUS 2112
Db 2310 GCCCGCTGAAGCTTGGCGGGAGGCGGCCACTCCACACCTCGCGCGCTCCCTGAGA 2251
Qy 2112 erGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLYsGlyHisGlnArgV 2132
Db 2250 GCCAGGCTCTGTCAGACCCGCTGCTCCAGACCCCGAGGGGTCAAAAGGTCAACACGGG 2191
Qy 2132 alValThrIleuAlaGlnHisIleSerGlyValIleThrGlnAspTYrThrArgHisAsP 2152
Db 2190 TGGTCACTGCGCCAGCAATCAAGAGGTCAACAAGACTACACCGGACACACC 2131
Qy 2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTYrSerPheProGlyAlaSerCyAsP 2172
Db 2130 CACAGAGAGTCAAGCGAACCTGCGCCCGCTTACTCTTCCCTGGGGGCAAGCTGGCC 2071
Qy 2172 iValLeuAspLeuArgArgProProSerAspLeuTYrLeuProProProAspHisGlyA 2192
Db 2070 CCGTCTCGAAGCTCCCGCCCGCCACCAAGTGAAGCTTACCTCCGCGCCCGGAGCAATGTG 2011
Qy 2192 laProAlaArgGlySerProHisSerGlyGlyLYsArgSerProGluProAsnLYsT 2212
Db 2010 CCGCGGCGCGTGGCTCCCCCAGAGGAGGGGCAAGAGGTCTCCAGAGCCAAACAAG 1951
Qy 2212 hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2232
Db 1950 CGTGGGTCTTGGTGGTGGAGAGCGTATTGAACCTGTCTCCACCGGAGGCGCATGA 1891
Qy 2232 hrGluProGlyHisSerArgSerAlaValTYrProLeuLeuTYrArgAspGlyGluGlnT 2252
Db 1890 CGAGAGCCAGGGGCACTCCCGAGGTGTGTGAACCGCTGTGTACCGGGAATGGGAAACA 1831
Qy 2252 hrGluProSerArgMetGlySerLYsSerProGlyAsnThrSerGlnProProAlaPheP 2272
Db 1830 CGAGGCCAGAGAGAGGTGCTCCAAAGTCTCCAGGCAACACAGCCAGCGCCAGCTTCT 1771

Qy 2272 heSerLYsLeuThrGluSerAsnSerAlaMetValLYsSerLYsLYsGlnIuIleAsnL 2292
Db 1770 TCAGCAAGCTGACCGAGAGCACTCCGCCATGTCAAGTCCAAAGAGATCAACA 1711
Qy 2292 YsLYsLeuAsnThrHisAsnArgAsnGluProGluTYrAsnIleSerGlnProGlyThG 2312
Db 1710 AGAAGCTGAACAACCAACCGGAATGAGCTGAAATACATATCAAGCCAGCTGGAGCG 1651
Qy 2312 luIlePheAsnMetProAlaIleThrGlyTYrGlyLeuMetThrTYrArgSerGlnAlaV 2332
Db 1650 AGATCTTCAATATGCTCCGCCCATCACCGGAACAGGCTTATGACTATTAACCCAGCGG 1591
Qy 2332 alGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLYsAlaLeuMetG 2352
Db 1590 TCGAGGAACATGCGAGCAACATAGGGCTGAGGGCCATTAATTGAAGGCACTCATG 1531
Qy 2352 lYLYsTYrAspGlnIuIuTPGluGluSerProProLeuSerLYsAsnAlaPheAsnProLeu 2372
Db 1530 GTAAATATGACCAAGTGGAAAGTCCCGCGCTCGATGCGCCAAATGCTTTAAACCTCTGA 1471
Qy 2372 snAlaSerAlaSerLeuProAlaIleMetProIleThrAlaAlaAspGlyArgSerAspH 2392
Db 1470 ATGCCAGTCCAGCTGCGCGCTGCTATGCTCATTAACGCTGCTGACGAGCGAGTGAAC 1411
Qy 2392 laThrLeuThrSerProGlyGlyGlyLYsValAlaLYsValSerGlyArgProSerSerA 2412
Db 1410 ACACACTCACTCGCAGGTGCGCGGAGAGCCCAAGGTCTCTGGCAGACCCAGACGCC 1351
Qy 2412 rGlyValAlaLYsSerProAlaProGlyLeuAlaSerGlyYAspArgProProSerValSerS 2432
Db 1350 GAAAGCCAAAGTCCCGGCGCGGCTGCGATCTGGGAGCCGGCAACCTTGTCTCT 1291
Qy 2432 erValHisSerGluGlyAspCyAsnArgArgThrProLeuThrAsnArgValITPGluA 2452
Db 1290 CAGTGCACCTCGAGGAGAGACTCAACCGCGGAGCGGCTTCAACCGCGTGGAGAG 1231
Qy 2452 sPArgProSerSerAlaGlySerThrProPheProTYrAsnProLeuIleMetArgLeuG 2472
Db 1230 ACAGGCCCTCGTCCGAGGTTCCAGCGCAATCCCTACAAACCCCTGATCATGCGGCTGC 1171
Qy 2472 luAlaGlyValMetAlaSerProProProProProGlyLeuProAlaGlySerGlyProLeuA 2492
Db 1170 AGCGGGGTCTATGGCTTCCCAACCCCGAGGCTCCCGGGGACGCGGCGCTTGC 1111
Qy 2492 laGlyProHisHisAlaTPAspGluGluProLYsProLeuLeuCySerGlnTYrGluT 2512
Db 1110 CTGGCCCCCAACAGGCTGGAGAGAGCCCAAGCACTGCTGTGCTGCAGTACGAGA 1051
Qy 2512 hrLeuSerAspSerGlu 2517
Db 1050 CACTCTCCGACAGCGAG 1034

Search completed: November 11, 2005, 10:58:53
Job time : 4786 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 11, 2005, 09:39:08 ; Search time 5611 Seconds
(without alignments)
3709.701 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQTWRATEPRYP.....WDEPRKLLCSQYETLSDSR 2517

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Deloxt 7.0

Searched: 9794790 seqs, 4134903567 residues
Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.p2n.model -DEV=xlp
-Q=/cgn2_1/USFTO.epool_p/US09522753/runat_02112005.173641_20852/app_query.fasta-1.2695
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MIMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.csl -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09522753 @CGN 1.1 1531 @runat_02112005.173641_20852
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FOGAPOP=6 -FOGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELOXT=7

Database :
Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10A_NEW_PUBCOMB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US10A_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUBCOMB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13215	100.0	7554	18 US-10-414-692-14	Sequence 14, Appl
2	13215	100.0	8561	18 US-10-174-014-11	Sequence 11, Appl
3	13215	100.0	8561	26 US-11-052-344-1	Sequence 1, Appl
4	13145	99.5	8667	14 US-10-087-192-653	Sequence 653, App
5	12978	98.2	7521	10 US-09-819-104A-3	Sequence 3, Appl
6	12978	98.2	8666	10 US-09-819-104A-1	Sequence 1, Appl
7	12978	98.2	8666	18 US-10-174-014-4	Sequence 4, Appl
8	12978	98.2	8666	21 US-10-723-860-1739	Sequence 1739, Ap
9	12845.5	97.2	9079	21 US-10-723-860-6114	Sequence 6114, Ap
10	10832.5	82.0	7386	10 US-09-819-104A-6	Sequence 6, Appl
11	10832.5	82.0	8544	10 US-09-819-104A-4	Sequence 4, Appl
12	10832.5	82.0	8544	22 US-10-764-420-1526	Sequence 1526, Ap
13	9691.5	73.3	7534	14 US-10-087-192-650	Sequence 650, App
14	4612	34.9	2930	16 US-10-146-473-14	Sequence 14, Appl
15	4612	34.9	2930	18 US-10-174-014-13	Sequence 13, Appl
16	4165.5	31.5	7940	18 US-10-341-434-235	Sequence 235, App
17	4165.5	31.5	7940	22 US-10-887-553A-664	Sequence 664, App
18	4147.5	31.4	7940	21 US-10-473-974-191	Sequence 191, App
19	4147.5	31.4	7940	22 US-10-956-157-1580	Sequence 1580, Ap
20	4147.5	31.4	7940	22 US-10-745-237-221	Sequence 221, App
21	3649	27.6	91141	14 US-10-087-192-649	Sequence 649, App
22	2951	22.3	23380	14 US-10-087-192-652	Sequence 652, App
23	2184	16.5	2745	18 US-10-414-692-15	Sequence 15, Appl
24	2133.5	16.1	22100	18 US-10-174-014-12	Sequence 12, Appl
25	1083	8.2	10910	24 US-10-745-237-23	Sequence 23, Appl
26	1083	8.2	10910	26 US-11-097-143-1937	Sequence 1937, Ap
27	1076	8.1	872	24 US-10-779-543-8757	Sequence 8757, Ap
28	880	6.7	718	22 US-10-278-698-515	Sequence 515, App
29	880	6.7	718	22 US-10-278-698-1031	Sequence 1031, Ap
30	746.5	5.6	520	9 US-09-920-300A-485	Sequence 485, App
31	746.5	5.6	520	14 US-10-033-528-485	Sequence 485, App
32	746.5	5.6	520	17 US-10-099-926-485	Sequence 485, App
33	746.5	5.6	520	24 US-10-961-527-485	Sequence 485, App
34	730.5	5.5	29467	22 US-10-937-730A-1	Sequence 1, Appl
35	730.5	5.5	67167	22 US-10-937-730A-3	Sequence 3, Appl
36	730.5	5.5	902568	16 US-10-156-761-1	Sequence 1, Appl
37	721	5.5	88400	22 US-10-844-716-1	Sequence 1, Appl
38	716.5	5.4	27705	10 US-09-942-025-12	Sequence 12, Appl
39	716.5	5.4	67311	10 US-09-942-025-11	Sequence 11, Appl
40	700	5.3	13416	16 US-10-205-032-11	Sequence 11, Appl
41	700	5.3	60196	16 US-10-205-032-1	Sequence 1, Appl
42	689.5	5.2	64492	19 US-10-378-083-1	Sequence 1, Appl
43	689	5.2	427	24 US-10-779-543-1895	Sequence 12895, A
44	685	5.2	30690	20 US-10-204-862A-1	Sequence 1, Appl
45	685	5.2	30690	26 US-11-005-196-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-414-692-14
; Sequence 14, Application US/10414692
; Publication No. US20030228607A1
; GENERAL INFORMATION:
; APPLICANT: X-Ceptor Therapeutics, Inc.
; TITLE OF INVENTION: Screening method and modulators having an improved therapeutic
; FILE REFERENCES: 8012-002-US
; CURRENT APPLICATION NUMBER: US/10/414,692
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 60/372,650
; NUMBER OF SEQ ID NOS: 86

; SOFTWARE: Patentin version 3.2
; SEQ ID NO 14
; LENGTH: 7554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-414-692-14

Alignment Scores:

Pred. No.:	0	Length:	7554
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	18	Gaps:	0

US-09-522-753-5 (1-2517) x US-10-414-692-14 (1-7554)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPdArgAlaThrGluProArgTyrPro 20
DB 1 ATGTGGGGCTCCACAGCTGTGGGCAAGCGTGGAGGGCCACTGAGCCCCGCTACCCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 61 CCCACAGCCTTTCCTACCCAGTGCAGATCGCCGAGCGCACAGACGTCGGGCTCTCG 120
QY 41 GlnTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACAGACACACTCCGCGCATGTGCTCCACCTGTGCGCGGCTCCATCATCCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 181 CCCACGGCGGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGAGATGAACGGTCCAG 240
QY 81 GlnLeuHisLeuArgProGlnSerHisSerTyrLeuProGlyLeuGlyLysSerGlnMet 100
DB 241 GAGCTTCACCTCGGCGCAAGTCCCACTATACCTCCGAGCTGGGGAAGTCAGAGATG 300
QY 101 GlnPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAAGAGCTGCTGCTGACCCCTGCTGCGACCG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspAspSer 140
DB 361 TCACCCCTGCTGCGCCACGGGCGAGCTGCGGAGTCTGAAGACCTCCACCAAGACCGCTAAC 420
QY 141 LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGluLeu 160
DB 421 CTGACGGGCAAGCTGGAAACGGTGTCTCCCCAGCCGCCCGACACTGACCTTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerLysGlnGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGTCGCGCAAGCTGTCTCAAGAGAGAGCTGATCCAGAACATGGAACCGCGTGAGC 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerLysLeuLysLysLysGlnGlnLeu 200
DB 541 CAGAGATCAACATGTGTAGAGCAGCAGATCTTAACTGAAAGAAAGAGCAGCAACGCTG 600
QY 201 GlnGlnGlnAlaAlaLysProProGlnProGlnLysProValSerProProIleGln 220
DB 601 GAGGAGAGAGGCTGCCAAGCCGCCGAGCTGAGAAACCCGTGTACCGCGCCCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlnAla 240
DB 661 TCGAAGACACCGCAGCTGTGTGAGATCATCTAGCAGAGAACCGGAAGAGGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnValGlnLeuProLeuTyrArgGlnPro 260
DB 721 GGACATCGGATTCGTGAAGAGCTCGGGGCCCGCAGGTGAGACTGCGCTGTGAACAACGAGCC 780
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TTCGACACCCCGGAGTATCATGAGAACATCAAAATTAACAGGCGGATGCGGAAGAGCTA 840

QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnThrLysGlnLysPheCysGln 300
DB 841 ATCTTGTACTTCANAGAGGAGGAATCAGCTCCGAAACATAGAGAGAGATTCCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysValGlnArgIleGlnAsnAsnPro 320
DB 901 CCTATGACCACTGATGAGGCTTGGAAAAAGGTGAGCGCATCCAAAAACACCCG 960
QY 321 ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle 340
DB 961 CCCCCGGCGCCAGAGAGAGCAAGGTGGCGGATCTCAAGAAAAGAGTTCCTCGATC 1020
QY 341 ArgLysGlnArgGlnLeuGlnGlnLysMetGlnSerArgValGlyGlnArgLysSerGly 360
DB 1021 CCGAAGCAGCGGACTGTGAGAGGCGCATGCGAGAGGCTGGCGCAGCGGCGGAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTGGGCGCCCGCAGCGAGACAGAGGTTCAGAGATCATCGATGGCTCTCA 1140
QY 381 GlnGlnGlnLysAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1141 GAGCAGAGAACTGGAGAAAGCAGATGGCGCAGCTGGCCGTGATCCCGCCATGCTGAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1201 GACGCTGACACAGCGCATCAAGTTCAATCAATGAAGGGCTTAATGGCCACCCCATG 1260
QY 421 LysValIleTyrLysAspArgGlnValMetAsnMetTPSerGlnGlnGlnLysGlnThrPhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCTATGAACATGTGAGTGAACAGAGAAAGAGACTTC 1320
QY 441 ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CGGAGAAAGTTGATGAGATCCCAAGAACTTGGCGTATGATGCATCTCCGAGAGAG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnLysTyrLys 480
DB 1381 AAGACAGGTGCTGAGTGCCTCTCTTAATTACTGACCTGAACAAAGATGAAGACTTAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgLysSerGlnGlnGlnGlnGln 500
DB 1441 AGCTGTGTAGACGGAGCTATGTGGCGCCCGGCGAAGCAGCAGCAGCAACAGCAGCAG 1500
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnLys 520
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCATGCCCGCAGCAGCAGAGAGGAGAA 1560
QY 521 AspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysProGlnValGlnAsn 540
DB 1561 GATGAGAAAGAGAAAGAAAGAGGCGGAGAGAGAGAGAAACCCGAGGTGGAGAAC 1620
QY 541 AspLysGlnAspLeuLeuLysGlnLysThrAspAspThrSerGlyGlnAspAsnAspGln 560
DB 1621 GACAAAGAAAGACTCTCTCAAGAGAAAGACAGACACTTCAGGGAGAGCAACAGCAG 1680
QY 561 LysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleYArgArgLysGly 580
DB 1681 AAGAGGCTGTGGCTCCAAAGCGCGCAAAACTGCCAAACAGCAGGAGAGAGCAAAAGGC 1740
QY 581 ArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnGlnAlaIleThrProGlnGln 600
DB 1741 CCGATCAACCCGCTCAATGCTATATGAGGCAACAGAGAGAGGCAATCAACCCCGCAGAG 1800
QY 601 SerAlaGlnLeuAlaSerMetGlnLeuAsnGlnLysSerArgTyrThrGlnGlnGlnMet 620
DB 1801 AGCGCGAGCTGCTCTCATGTAGAGCTGAATGAGATTCTCGCTGACACAAAGAGAAATG 1860
QY 621 GlnThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTyrPheAlaIleAlaArgMet 640
DB 1861 GAAACGCGCAAGAAAGTCTCTTGGAACAGCGCCGCAATGTGGCGCATCGCCCGGATG 1920
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660

1921 GTGGGCTCCAAAGCTGTGTCCAGGTAAAGAACTTCTACTTCAACTAACAAAGAGCGAG 1980
Qy AsnLeuAspGluIleLeuGlnGlnHisIleIysLeuIysMetGluIlyrGluIyrAsnAlaArg 680
Qy 661 AACCTCATGTAGTGAATCTTGACAGACACAAAGCTTAAGAGATGAGAGAGAGAGAGAGCGCGG 2040
Db 1981 AACCTCATGTAGTGAATCTTGACAGACACAAAGCTTAAGAGATGAGAGAGAGAGAGAGCGCGG 2040
Qy ArgIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 700
Db 681 AGGAAGAAAGAGAAAGCGG 2100
Qy 2041 AGGAAGAAAGAGAAAGCGG 2100
Qy GlnAspGlnGlnMetGlnIleSerGlyIleSerGlyIleSerGlyIleSerGlyIleSerGlyIle 720
Db 701 GAGATAGAGAGATGAGAGCGCTCGCGCGCTGAGCGGAAATGAGAGAGAGATGAGAGAGAG 2160
Qy 721 AAGGATAGAGAGATGAGAGCGCTCGCGCGCTGAGCGGAAATGAGAGAGAGATGAGAGAGAG 2160
Db 2161 GCTAAAGCTTAACTGCTCTGAGAAATGAGGTGCGCGAGGGGAAATGAGAGAGAGCGCGCGG 2220
Qy ThrValAsnAsnSerSerAspThrGlnIleProSerProHisIleThrGlnIleAlaIlyr 760
Db 741 ACTGTCAACAACACACTGACAGACCGAGAGCATCCCTCTCTCTCAACATGAGCGCGCGAG 2280
Qy 2221 ACTGTCAACAACACACTGACAGACCGAGAGCATCCCTCTCTCTCTCAACATGAGCGCGCGAG 2280
Qy AspThrGlnIleAsnGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 780
Db 761 GACACAGGCGAGAAATGGGCGG 2340
Qy 781 G1ProProthProProProArgThrSerArgAlaProIleGluProthProthProAlaSer 800
Db 2341 GGGCCACCCACCCACCGAGAGACATCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 2400
Qy 801 GlnAlaThrGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 820
Db 2401 GAAGCCACCGGAGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 2460
Qy 821 ValIleIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 840
Db 2461 GTGTGTCCCAAGG 2520
Qy 841 GlnGlnGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 860
Db 2521 GAGG 2580
Qy 861 ProValIlyrSerGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 880
Db 2581 CCGGTCAAG 2640
Qy 881 GlnAlaAlaGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 900
Db 2641 GAGGCGCGCTGAGGCGAGCGCGCGAGGCGCGCTCAAGGAGAGAGAGAGAGAGAGAGAGG 2700
Qy 901 GlnAlaThrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 920
Db 2701 GGGAGGGCGCGAGG 2760
Qy 921 CysSerIleAspGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 940
Db 2761 TGGAGTGCACACAGAGTGAAGG 2820
Qy 941 ArgProSerIleuIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 960
Db 2821 AGGCGCGCGCT 2880
Qy 961 LeuAspIleuIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 980
Db 2881 CTGAGCTTGAAGG 2940
Qy 981 ValHisGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1000
Db 2941 GTCCATGAGCGCGCGGAGG 3000
Qy 1001 ProProGlnIleuIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1020

3001 CCACCGAAAACTTGACCGGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3060
Qy 1021 G1IlyrSerIlyrSerProAlaProProAlaAspIlyrGlnIlyrIlyrIlyrIlyrIlyrIlyr 1040
Db 3061 GGCAGG 3120
Qy 1041 IlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1060
Db 3121 AAGCTGCTGGGAGG 3180
Qy 1061 GlnValIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1080
Db 3181 GAGGTGATCAAGG 3240
Qy 1081 G1IlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1100
Db 3241 GGTACCCACTGCGCGCTTGGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 3300
Qy 1101 ThrIleSerAsnProProProIleuIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1120
Db 3301 ACCATCTCCAACCGCGCTCCCTCATCTCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGG 3360
Qy 1121 GlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1140
Db 3361 CAATAGTGCATCTCCCAAGG 3420
Qy 1141 AlaIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1160
Db 3421 GCGAGG 3480
Qy 1161 LeuAlaProPheSerGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1180
Db 3481 CTGGCACTCTTCAAGCGAGG 3540
Qy 1181 ProGlnSerIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1200
Db 3541 CCGGAGG 3600
Qy 1201 G1IlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1220
Db 3601 GGTCTGATTCGGAGCGAGG 3660
Qy 1221 SerAlaIleThrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1240
Db 3661 AGGCGCATCATACAGCGCGCTCATACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 3720
Qy 1241 G1IlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1260
Db 3721 GGCACCATCACAGAGATCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 3780
Qy 1261 AspSerIleuProIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1280
Db 3781 GACAGCTGCGCAAGGAGG 3840
Qy 1281 GlnGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1300
Db 3841 GAGGTGCGATGCTGTGAGG 3900
Qy 1301 ProHisGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1320
Db 3901 CCCCATGAGG 3960
Qy 1321 AlaIleSerSerIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1340
Db 3961 GCCATCTCTCAAGG 4020
Qy 1341 SerProHisIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1360
Db 4021 AGGCGCGAGG 4080
Qy 1361 ArgSerIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr 1380
Db 4081 CGGTCTTACGTGAGG 4140

QY 1381 GIUGIYThrProProProProSerArgAspLeuThrGluAlaTyrIleThrGln 1400
DB 4141 GAGGGACGCTCCGCGCCACCGCCCTCACGGGACCTGACCGAGGCTTACAAGACGAC 4200
QY 1401 AlaleuGIYProLeuIleLeuIleProAlaHisGluGluValAlaThrValIleGlu 1420
DB 4201 GCCCTGGGGCCCCCTGAAGCTGAAGCGGGCCCATGAAGGCGCTGGTGGCCACGGGGAAGAG 4260
QY 1421 AlaGIYArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
DB 4261 GGGGGCGCTCCATCATAGATCCGCGCGAGAGCTCGGACACCGCCGAGCTGCCCC 4320
QY 1441 LeuAlaProArgProLeuIleGluGluSerIleThrGlnGluYThrProLeuIleTyrAsp 1460
DB 4321 CTGGCCCCCGCGCGCTCAAGAGGGGCTCCATCACCGAGGACCGCCGCTCAAGTACGAC 4380
QY 1461 ThrGluAlaSerThrThrGluSerIleValHisAspValArgSerLeuIleGlySerPro 1480
DB 4381 ACCGGCGCTCCACCTGGCTTCCAAAAGACGAGTACGCTCTCATCGGACGCCCC 4440
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGln 1500
DB 4441 GGGCGGACGTTCCACCGCTGACCGCGTGGATGTATGGCCGACCGCCGGGACCTGGAA 4500
QY 1501 ArgAlaCysTyrGluGluSerLeuIleSerArgProGluYThrAlaSerSerGlyGly 1520
DB 4501 CGTGCTGTACGAGGAGAGCTGAAGAGCGGGCCAGGACCGCCGACGACTGGGGGGC 4560
QY 1521 SerIleAlaArgGluValaProValIleValProGluLeuGluYIleProArgGlnSerPro 1540
DB 4561 TCCATTGGCCGGCGCGCCCGGCTATTGTGCTGACCTGGTAAAGCGCGGACGACCCCC 4620
QY 1541 LeuThrTyrGluAspHisGluAlaProPheAlaGluHisLeuProArgGlySerProVal 1560
DB 4621 CTGACCTATGAGGACCAACGGGGGACCTTTGCCGGGACCTCCCAAGAGTTGCCCGG 4680
QY 1561 ThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerIleYsaIaSer 1580
DB 4681 ACCATCGGGAGCGCACCGCGCGCTGACGAGAGGACGCTTTCTGTCAGCAAGGATCC 4740
QY 1581 GlnAspArgIleLeuThrSerThrProArgGluIleAlaIleSerProHisSerThrVal 1600
DB 4741 CAGGACCGGAAGCTGACGTGAGGCTCGTGAATGCGCAAGTCCCGGACAGCACCGG 4800
QY 1601 ProGluHisHisProHisProIleSerProTyrGluHisIleLeuLeuArgGluValSerGly 1620
DB 4801 CCGGAGACCAACCAACCCCATCTGCGCTATGAGCACTGCTGGGGCGTGAATGGC 4860
QY 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
DB 4861 GTGGACCTGATCGACGACATCCCGCTGGCTTGCACCCACCTCATACCCCGCGGC 4920
QY 1641 IleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisIleLeuAlaProAsnProThr 1660
DB 4921 ATCCCTCTGAGCGACCGCTGCTACTACTGCGCCGACACTGCGCCCAACCCACCC 4980
QY 1661 TyrProHisLeuTyrProProTyrLeuIleArgGluTyrProAspThrAlaAlaLeuGln 1680
DB 4981 TACCGGACCTGTACCAACCTTACTCTACCTCGGGCTTACCCGACACGGCGGGCTGGAG 5040
QY 1681 AsnArgGlnThrIleIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAla 1700
DB 5041 AACCGGACACCATCATGACTACATCACTTCGACAGAGATGACCAACAACAGCGGC 5100
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGluLeuSerProArgGluSerSerLeu 1720
DB 5101 ACCGCGCATGGCCGAGGAGCTGATATGCTGAAGGGGCTCTCGCCCGGACGTCCTGCTG 5160
QY 1721 AlaLeuAsnTyrAlaAlaGluProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
DB 5161 GCACTAACTACCTCGGGGTCCCGAGGACATGACCTGTGCCAAGTGCACACCTG 5220

QY 1741 ProValLeuValProProThrProGluYThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
DB 5221 CCTGTGCTGTGTCGCCCGCACACAGGACCCGACGACCGCATGAGCGCTTGTCTAC 5280
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
DB 5281 CTCCCAACCGCGCCCGAGCCCTTACAGACCGCCACAGAGCTCCCACTCTCCAGGA 5340
QY 1781 GlyProThrHisLeuThrIleYsaProThrThrThrSerSerSerGluArgGluAspArg 1800
DB 5341 GGTCCAAACACCTTGACAAAACCAACACACAGTCTCTGTCGAGGGGAGGAGCGCCG 5400
QY 1801 AspArgGluArgAspArgAspArgGluArgGluYSerIleLeuThrSerThrThrThr 1820
DB 5401 GATCAGAGCGGGAGCGGGATGGAGCGGAAAGTTCATCTCACTTCCACACAGAGC 5460
QY 1821 ValGluHisAlaProIleTyrArgProGluYThrGlnGlnSerSerGlySerSerGlySer 1840
DB 5461 GTGGAGCAGCACCATCTGGAGACTGTTCAGAGCCCTGACAGACAGCGGACGCGGACG 5520
QY 1841 SerGlyGlyGlyGlyYserSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5521 AGCGCGGGGGTGGGGGACAGACGCGCGCCCTCCATCCCATCCACACGACGAC 5580
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5581 TCGCCATCTCTCCCTGGACCCGAGATGCTTCCAGACAGACCCAGTGTGCTTCAAC 5640
QY 1881 ThrGlyMetIleGlyIleIleThrAlaValGluProSerIleYsaProThrValLeuArgSer 1900
DB 5641 ACGGCAATGAAGGATATCATCACCTGTGGAGCCACGAAACCCAGCGTCTCGAGTCC 5700
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5701 ACCTCACTCTCTCAACCGCTGCGCAGCTGCACATTCCTCCACTGCGACCACTGCCA 5760
QY 1921 LeuGlyGlyThrLeuAspGlyValTyrProThrIleLeuMetGluProValLeuLeuProIlys 1940
DB 5761 CTGGGGGGGACCTCGATGGGGTCTACCTTACCTCATGAGACCGCTTGTGCTGCCAAG 5820
QY 1941 GlnAlaProArgValAlaArgProGluArgProArgProAlaAspThrGluHisAlaPheLeu 1960
DB 5821 GAGGCGCCCGGGGTGCGCGGCGAGCGGCCCCGAGACGACACCGGCATGCTTCTC 5880
QY 1961 AlaYsaProProAlaArgSerGlyLeuGluProAlaSerSerProSerIleGlySerGln 1980
DB 5881 GCCAAGCCCCCAGCGCGCTCGGGCTGGAGCCGCTCTCCCGCCGACAAAGGCTCGGAG 5940
QY 1981 ProArgProLeuValProProValSerGluHisAlaThrIleAlaArgThrProAlaIlys 2000
DB 5941 CCGCGCCCTTGTAGTGTCTCTGTGCGCACGGCCACCATGCGCCGACCCCTGGAAAG 6000
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
DB 6001 AACCTGCACTTCAACGACCGGACCGGACCGCGCGGCCACGTGCGCTCGGCTCGGAC 6060
QY 2021 ProHisArgGluYsaThrGlnSerIleYsaProPheSerIleGlnGluLeuLeuArgSer 2040
DB 6061 CCGCACCGGGAAGAAGCTCAAGTAAACCTTTTCATGACAGAACTGGAACTCCCTTGT 6120
QY 2041 LeuGlyTyrHisGlySerSerTyrSerProGluGluValGluProValSerProValSer 2060
DB 6121 CTGGGTTTACACGCGGACGAGCTTACAGCCCCGAAAGGGTGAAGCCCTGACCCCTGGAGC 6180
QY 2061 SerProSerLeuThrHisAspIleYsaGlyLeuProIlysHisIleGluGluLeuAspIleSer 2080
DB 6181 TACACCAAGTGTACCCACGACAAAGGGGCTCCCAAGACCTGGAAGAGCTCCAGAAAGC 6240
QY 2081 HisLeuGluGlyGluLeuArgProIlysGlnProGlyProValIleLeuGluGlyGluAla 2100
DB 6241 CACTGGAAGGGAGGAGCTGGCGCCAAAGCAGGAGGCCCGGTGAAGCTTGGCGGGAGGCGC 6300
QY 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120

Db 6301 GCCACCTCCACACCTGGCGGCGCTGCTGAGAGCCAGCCCTGCTCCAGCCCTGCTC 6360
Qy 2121 GlnThrAlaProGlyValIleGlyGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
Db 6361 CAGACCGGCCCCAGGGGCTCAAGGTCACAGCGGGGGTGGTCACTGGCCCAACACATCATGT 6420
Qy 2141 GluValIleThrGlnAspTyrThrArgHisHisIleProGlnGlnLeuSerAlaProLeuPro 2160
Db 6421 GAGGTATCACACAGAGCATACACCGGACACCAACAGAGCTCAAGGCACTCCCTGCCC 6480
Qy 2161 AlaProLeuTyrSerPheProGlyAlaSerCyAspProValLeuAspLeuArgArgProPro 2180
Db 6481 GCCCCCTCTACTCTCTCCCTGGGGGAGCTGCGCCCTGCTGAGACTCCGCGGCCACACC 6540
Qy 2181 SerAspLeuTyrLeuProProProAspHisGlyValAlaProAlaArgGlySerProHisSer 2200
Db 6541 AGTGACCTCTACCTCCGCGCCCGGACCAATGATGCCCGGCTGTGCTCCCTCCACAGC 6600
Qy 2201 GlnGlyGlyIleAspArgSerProGluProAsnIleThrSerValIleGlyGlyGlyAsp 2220
Db 6601 GAAGGGGGCAAGAGTCTCCAGAGCCAAACAGACGTGGTCTGGGTGGTGGTGGAGAC 6660
Qy 2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
Db 6661 GGTATTGAACTGTGTCCCAACCGAGGATGACGAGCAGGACACTCCCGGAGTGCT 6720
Qy 2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerIle 2260
Db 6721 GTGTACCGGTGCTGTAACCGAGATGGGAAACAGCGAGGCCAGAGATGGGTCTCAAG 6780
Qy 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerIleThrGluSerAsnSer 2280
Db 6781 TCTCCAGGCAACACAGCCAGCCCGCTTCTTCAGCAAGCTGACCGAGAGCACTCC 6840
Qy 2281 AlaMetValIleSerIleValGlnGluIleAsnIleValLeuAsnThrHisAsnArgAsn 2300
Db 6841 GCCATGTCTCAAGTCAAGAGCAAGAGATCAACAGAGAGTCAACACCAACCGGAT 6900
Qy 2301 GluProGluIleThrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr 2320
Db 6901 GAGCCTGAATACATATACAGCCAGCTGAGAGGATCTTCAATATAGCCCGCATCAC 6960
Qy 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisIleSerThrAsnMet 2340
Db 6961 GGAACAGGCTTATGACTTATAGACGACGCGGTGAGAACATGCGACACCAACATG 7020
Qy 2341 GlyLeuGlnAlaIleIleArgIleValIleLeuMetGlyIleTyrArgGlnTyrGluIleSer 2360
Db 7021 GGGCTGAGAGGCTAATTAAGAAAGGCACTATGGGTAAATATGACCACTGGGAAAGTCC 7080
Qy 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
Db 7081 CCGCCCTCTACGCGCAATGCTTTTAAACCTCTGAATCCAGTCCAGCTCCCGCTGCT 7140
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrIleThrSerProGlyGlyGly 2400
Db 7141 ATGCCATTAACCTGCTGACGAGAGGATGACACACACTTCACTGCGAGGTGGGCT 7200
Qy 2401 GlyIleValAlaValSerGlyArgProSerSerArgIleValIleIleSerProAlaProGly 2420
Db 7201 GGGAAAGGCCAAGGCTCTGGCAGACCCAGAGCGGAAAGCCAAAGTCCCGGCGCGG 7260
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCyAsn 2440
Db 7261 CTGGCACTCTGGGAGACCGGCACTCTGTCTCTCAAGTGCATCGAGAGGAGCTCCAAC 7320
Qy 2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
Db 7321 CGCGGAGCGGCTCAACACCGCGTGTGGAGAGACAGGCTCTGTCGAGGTTCCAGC 7380
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyAlaMetAlaSerProPro 2480

Db 7381 CCATTCCCTCAACACCCCTGATCATGCGGTGACAGCGGGGTGCATGGCTTCCCAACC 7440
Qy 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisIleAlaThrAspGlu 2500
Db 7441 CCACGGGCGCTCCCGGCGGAGCGGGCCCTGCTGAGCCCCCAACAGCTGGGAGCAG 7500
Qy 2501 GluProIleProLeuLeuCySerGlnTyrGluThrIleSerAspSerGlu 2517
Db 7501 GAGCCCAAGCACTGCTGCTGCTGCACTACGAGACACTTCCGACAGCAG 7551

RESULT 2
US-10-174-014-11
; Sequence 11, Application US/10174014
; Publication No. US20040005292A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freiler
; APPLICANT: Kenneth W. Doble
; TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
; FILE REFERENCE: PTS-0012
; CURRENT APPLICATION NUMBER: US/10/174,014
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 73
; SEQ ID NO 11
; LENGTH: 8561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2) ... (7555)
US-10-174-014-11

Alignment Scores:
Pred. No.: 0 Length: 8561
Score: 13215.00 Matches: 2517
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-09-522-753-5 (1-2517) x US-10-174-014-11 (1-8561)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro 20
Db 2 ATGTCCGGGCTTCCACACAGCTTGTGGACACAGCTGAGGGCCACTGAGCCCTTACCG 61
Qy 21 ProHisSerIleuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 62 CCCACAGCCTTCTTACCAAGTGCAGATGCGCGGACGACACAGCGAGTGGGCTCTG 121
Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 122 GAGTACACAGACACATCCCGGCACTATGCTCCACACTGTGCGGGGCTTCATCATCCAG 181
Qy 61 ProGlnArgArgArgProSerIleuSerGluPheGlnProGlyIleGlnIleValGlySerGln 80
Db 182 CCCACAGGCGAGGCGCTCTCTGCTGCTGAGTTCAGAGCCGAGATGAACGATCCAG 241
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleValSerGluMet 100
Db 242 GAGCTCACCTGGGCGCAGAGTCCCACTATACCTGCCGAGTGGGAAAGTCAAGATG 301
Qy 101 GluPheIleGluSerIleAspArgProAlaGluLeuLeuLeuProAspProLeuLeuArgPro 120
Db 302 GAGTTCATTGAAGCAAGAGCGCTCGCTAGAGCTGCTCCTGAGACCCCTGCTGCAGCG 361
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlyAspLeuThrIleValAspArgSer 140
Db 362 TACACCTCTGCGGCGACGCGGCGAGCTGCGGATTCGAAAGACTTCAACAGACCTTACG 421
Qy 141 LeuThrGlyIleValLeuGluProValSerProProSerProProHisIleThrAspProGluLeu 160
Db 422 CTGACGCGGAGAGCTGGAACGGGTGTCTCCCGACAGCCCGCCGACACTGACCTGAGCTG 481

QY 161 GIuleuValPProProArgLeuSerLySGluLeuIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGTGCCCCCGCCAGCTGTCCAAAGAGAGCTGATCCAGAACATGAGACCGCGTGGAC 541
QY 181 ArgGluIleThMeValGluGlnGlnIleSerLySLeuLySLeuLySGlnGlnIleu 200
DB 542 CAGAGATCATCATGTAGTACAGCAGATCTCTAACTCAAGAAAGAAAGCAGCAACGCTG 601
QY 201 GluGluGlnAlaAlaLySProProGluProGluLySProValSerProProIleGlu 220
DB 602 GAGAGAGAGCTGCCCCAGCCGCCAGCTGAGAACCCCGTGTACCCGCCCCCATCGAG 661
QY 221 SerLySLeuArgSerLeuValGlnIleIleTyArgSpGluAsnArgLySLeuValaGlu 240
DB 662 TCGAAGCACCCGACGCTGTGAGATCATCTAGCAGAGAACCGGAAGAAAGCTGAAGCT 721
QY 241 AlaHisArgIleLeuGlnGluLyLeuGlyProGlnValGluLeuProLeuTyArgGlnPro 260
DB 722 GCACATCGGATCTGGAGAGCCCTGGGGCCCGAGGTGAGCTGCCCTGTACAAACGAGCC 781
QY 261 SerAspThrArgGluThrHisGluAsnIleLySLeuGlnAlaMetArgLySLeu 280
DB 782 TCCGACACCCGCGAGTATCATAGAACATCAAAATTAACCGCGATGCCGAAGAGCTA 841
QY 281 IleLeuTyRPhelysArgArgAsnHisAlaArgLySGlnTrpLySGlnLySPhelysGln 300
DB 842 ATCTTGTACTTCAAGAGAGGAATCACGCTCGGAACAATGGAAGCAAGATTCTCCAG 901
QY 301 ArgTyArgSpGlnLeuMetGluAlaLeuGluLySLeuValGluArgIleGluAsnAsnPro 320
DB 902 CGCTAAGACAGCTCATGGAGGCTTGGAAGAAAAAGTGGAGCGCATCGAAAAACAACCG 961
QY 321 ArgArgArgAlaLySGlnSerLySValArgGluTyRTrpGluLySGlnPheProGluIle 340
DB 962 CGCGCGCGGCGCAAGAGAGCAAGGTGCGGAGTACTACGAAAAAGCATTCCTTGAGATC 1021
QY 341 ArgLySGlnArgGluLeuGlnGlnIleuArgMetGlnSerArgValGlyGlnArgLySGly 360
DB 1022 CCCAAGCACCGGAGCTGCAGAGAGCGCATGCAAGCAGGGTGGCCAGCGGCGCACTGGG 1081
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1082 CTGTCCATGTCCGGCCCCCGCAGCGAGCAAGAGTGTCAAGATCATCGATGGCTCTCA 1141
QY 381 GluGlnGlnAsnLeuGlnLySGlnMetArgGlnLeuAlaValIleProPheMetLeuTyR 400
DB 1142 GAGCAGGAGAACCTGGAGAGAGAGATGCGCAGCTGCGCGTGAATCCCGCCATGCTGAC 1201
QY 401 AspAlaAspGlnGlnArgIleLySPhelIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1202 GAGCGTCAGCAGAGCGGATCAAGTTCAATCAATGACGAGCTTATGGCCGACCCCATG 1261
QY 421 LySValTyRlyAspArgGlnValMetAsnMetTrpSerGluGlnGlnLySGluTrpPhe 440
DB 1262 AAGGTGTACAAGAGCCGCGAGTCAATGAACATGTGAGTGAAGAGAGAGAGACCTTC 1321
QY 441 ArgGlnLySPhelysMetGlnHisProLySAsnPhelysLeuIleAlaSerPheLeuGlnArg 460
DB 1322 CGGGAAGAGTTCATGAGCATCCCAAGAACCTTGGCTGATGCCATCATTCCTGGAGAGG 1381
QY 461 LySThrValAlaGluCySValLeuTyRlyTyRLeuThrLySLeuAsnGluAsnTyRlyS 480
DB 1382 AAGACAGTGGCTAGTGGCGTCTCTATTACTACTGACTAAGAAATGAGAACTATAAG 1441
QY 481 SerLeuValArgArgSerTyRArgArgArgGlyLySLeuGlnGlnGlnGlnGln 500
DB 1442 AGCTGTGTGAGAGAGGTATCCGCGCGCGCAAGAGCAGCAGCAACAACGACGACAG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561

QY 521 ArgGluLySGluLySGluLySGluAlaGluLySGluGluGluLySProGluValGluAsn 540
DB 1562 GATGAGAAAGAGAAAGAAAGAGCGCGAGAGAGAGAGAGAAACCGGAGGTGGAGAAC 1621
QY 541 AspLySGluAspLeuLeuLySGluLySThrAspAspThrSerGlyGluAspAsnAspGlu 560
DB 1622 GACAGAGAGACCTCTCAAGAGAGAGAGAGAGACACTTCAGGGAGAGAGACACAGCAG 1681
QY 561 LySGluAlaValAlaSerLySGlyArgLySThrAlaAsnSerGlnGlyArgArgLySly 580
DB 1682 AAGAGGCTGTGCTCCAAAGCCGCAAAATCGCCAAACAGCAGGAGAGAGCAAAAGGC 1741
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1742 CCGATCACCCCGCTCAATGAGCTATAGGCCCAACAGAGAGAGCCATCACCCCGCAGAG 1801
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlnGluGluMet 620
DB 1802 AGCGCCGAGCTGCTCTCCATGAGCTGAATGAGATTCTCGCTGACAGAGAGAGAAATG 1861
QY 621 GluThrAlaLySLeuGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
DB 1862 GAAACAGCCAGAAAGGTCTCTGGAACACGCGCCGCACTGGTCGGCCATCGCCCGAGATG 1921
QY 641 ValGlySerLySThrValSerGlnCySLeuAsnPhelysPheAsnTyRlySLeuArgGln 660
DB 1922 GTGGGCTCCAAAGACTGTGTGAGTGAAGACTTCTTACTTAACACAAAGAGAGCAG 1981
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLySLeuLySMetGluLySGluArgAsnAlaArg 680
DB 1982 AACCTCGATGAAATCTTGACAGACAGCAAGCTGMAAGATGGAAGAGAGAGAACGCGCG 2041
QY 681 ArgLySLeuLySLeuAlaProAlaAlaAsnSerGluGluAlaAlaPheProProVal 700
DB 2042 AAGAAAGAAAGAAAGCCCGCGCGCGGCGCACGAGAGAGCTGCAATCCCGCCGAGTG 2101
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
DB 2102 GAGAGTGAAGAGATGAGCGCTCGCGCGTGAACCGGAAGAGAGAGATGTGGAGAGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCySLeuSerGlyProAla 740
DB 2162 GCTGAAGCTTTCATGTCTCTGGAAATGAGGTGCCAGAGGGGAATGCAATGGCCCAAGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLyS 760
DB 2222 ACTGTCAACACAGCTCAAGACACGAGAGACATCCCTCTCTCAACATGAGGCGCCAG 2281
QY 761 AspThrGlyGlnAsnGlyProLySProProAlaThrLeuGlyAlaAspGlyProProPro 780
DB 2282 GACACAGGCGCAGAAATGGGCCCAAGCCCGCCAGCCACTTGCGCGCGCAGCGGCCACCCCA 2341
QY 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
DB 2342 GGCCCAACCCACACACAGAGAGACATCCCGGCCCATTTGAGGCCACCCCGGCTCT 2401
QY 801 GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
DB 2402 GAAGCCACCGAGGCCCTAGCGCCCAACAGACCCCATCGGCTGTGACCTCTCTCT 2461
QY 821 ValValProLySGluGluLySGluGluGluThrAlaAlaAlaProProValGluGluGly 840
DB 2462 GTGTGTCCCAAG 2521
QY 841 GluGluGlnLySProProAlaAlaGluGluLeuAlaValAspThrGlyLySLeuAlaGluGlu 860
DB 2522 GAGAGACAG 2581
QY 861 ProValLySLeuSerGluCyThrGlnGluGluAlaGluGluGlyProAlaLySGlyLySAspAla 880
DB 2582 CCGGTCAAGAGAGAGTGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2641
QY 881 GluAlaAlaGluAlaThrAlaGluGluAlaLeuLySLeuLySGluGluGlySer 900

Db 2642 GAGGCCCTGAGGCGCAGGCGGCGCTCAAGCAGAGAAAGAGGCGGAGC 2701
Qy 901 G1YArg1aThrThrAlaYsSerSerG1YAlaProGlnaPseArpSerSerAlaThr 920
Db 2702 GCGAGGCGCCACATTCGCAAGAGCTCGGGCGCCCCAGCAGCAGCTCCAGTGTACC 2761
Qy 921 CysSer1aAapG1YValaAapG1YAlaG1YG1YAp1YsAaAaG1YLeuLeuSerPro 940
Db 2762 TGCAGTGCAGACGAGGTGATGAGCGCGAGGCGGCGCAAGAACCGGCTGTCTCCCA 2821
Qy 941 ArgProSerLeuLeuThrProThrG1YAspProAlaAlaAaAaAspProGlnLYsPro 960
Db 2822 AGGCGCCAGCTCTCCACCCGACCTGGCGACCCCGGCGCAATGCTCAACCCCAAAACCA 2881
Qy 961 LeuAspLeuYsG1YLeuYsG1YAlaAlaAla1LeuProPoli1eGlnVal1ThLYs 980
Db 2882 CTGGACCTGAAGAGCTGAAGAGCAGGAGCGGCTGCATCCCGCCATCCAGTCAACAAA 2941
Qy 981 ValH1eG1YProProArG1YAspAlaAlaProThrYsProAlaProProAlaProPro 1000
Db 2942 GTCATGAGCGCCCGCGGAGAGCAGCTCCACCAAGCAGCTCCCGCAGCCCAACG 3001
Qy 1001 ProProG1YAsnLeuG1YProG1YAspAlaProGlnG1YSerSerProArg 1020
Db 3002 CCACCGCAAAACCTGCAGCGCGAGCAGCGCCCTCAGCAGCTCGGACAGCGCCCGG 3061
Qy 1021 G1YLYsSerArgSerProAlaProProAlaAspYsG1YAlaAlaG1YAlaG1Y 1040
Db 3062 GCGCAAGAGCAGAGCGCGGAGCGCCCGCGCAAGAGGCTTCGAGCGGAGCGCCAG 3121
Qy 1041 YsLeuProG1YAspProProCySTpThrSerG1YLeuProPheProValProProArg 1060
Db 3122 AACCTGCTGGGAGACCCCTTGTCTGATCTTCGGGCTGCTCCCGTGGCCCGCCCT 3181
Qy 1061 G1YVal1YleYsAlaSerProH1aAlaProAspProSerAlaPheSer1YAlaProPro 1080
Db 3182 GAGGTATCAAGGCTCCCGCATGCGCCGAGACCCCTCAGCTTCCTCAAGCTCCACT 3241
Qy 1081 G1YH1aProLeuProLeuG1YLeuH1aAspThrAlaArgProValLeuProArgProPro 1100
Db 3242 GGTACCCCATGCGCCCTGGGCTCTCAATGACATGCGCGCCGCTCGCGCGCCACCC 3301
Qy 1101 Thr1YLeSerAnpProProProLeuH1eSerSerAlaYsH1aPseArpVal1LeuG1YArg 1120
Db 3302 ACCATCTCCAAACCGGCTCCCTCATCTCTCTGCCAAGCAACCCAGCTCTCGAGAGG 3361
Qy 1121 Gln1YleG1YAla1YleSerGlnG1YMetSerValG1YLeuH1eValProTYrSerG1YH1e 1140
Db 3362 CAATAGGTGCCATCTCCCAAGAAATGTCGATCCAGCTCCAGCTCCCGTACTCAGAGCAT 3421
Qy 1141 AlaLYsAlaProAlaG1YProVal1ThrMetG1YLeuProLeuProMetArpProLYs 1160
Db 3422 GCCAAAGCGCCCGGTGGGCTCTGTACCATGGGGCTGCTGCCATGAGCCCAAAAG 3481
Qy 1161 LeuAlaProPheSerG1YValYsG1YGlnG1YLeuSerProArgG1YAlaG1YPro 1180
Db 3482 CTGGGACCTTCACGGAGAGAGAGCAGAGCACTGTCCCAAGGGCGCAGCTGGGCCA 3541
Qy 1181 ProG1YSerLeuG1YValProThrAlaGlnG1YAlaSerValLeuArgG1YThraAlaLeu 1200
Db 3542 CCGAAGAGCTGGGGTGGCCACAGCCAGAGAGGCTCGTGTGAGAGGAGCAGCTGTG 3601
Qy 1201 G1YSerValProG1YG1YSer1YleThrLYsG1Y1aProSerThrArgValProSerArp 1220
Db 3602 GGCTCATGTCGGGGCGAAGCATCACAAAGGCAATTCACAGCACAGGGTGGCTCGAGC 3661
Qy 1221 SerAla1YleThTYrArgG1YSer1YleThrH1eG1YThrProAlaArpValLeuTYrLYs 1240
Db 3662 AGGCGCATCAACACGGCGCTCATCCACAGGAGCGCAGCTGACGTCTGTACAG 3721
Qy 1241 G1YTh1YleThrArg1Yle1YleG1YAspSerProSerArgLeuArpArg1YArgG1Y 1260

Db 3722 GGCACATCAACAAGATCATGTGGCGAGAGACGCCGAGTGTGACCGCGCGGAG 3781
Qy 1261 AspSerLeuProLYsG1YH1eVal1YleTYrGlnG1YValYsG1YH1eValLeuSerTYr 1280
Db 3782 GACAGCTGGCCAAAGGCGCAGTATATCAGAGGCAAGAGGCGCAGCTTGTCTCAT 3841
Qy 1281 G1YG1YMeSerVal1ThrGlnCYsSerLYsG1YAspG1YArgSerSerSerG1YPro 1300
Db 3842 GAGGTGAGATGTCTGTGACCAAGTGTCCAGAGAGAGCAGAGACAGCTCAGAGACC 3901
Qy 1301 ProH1eG1YThraAlaAlaProLYsArgThTYrAspMetMetG1YArgValG1YArg 1320
Db 3902 CCCCATGAGACGGCGCGCCCAAGCGCATATGACATGATGAGGCGCGCTGTGGGAGA 3961
Qy 1321 Ala1YleSerSer1aSer1YleG1YLeuMetG1YArgAla1YleProProG1YArgH1e 1340
Db 3962 GCCATCTCTCAGCCAGCATGAAAGCTTATAGGCGGTCATCCCGCGAGCGACAC 4021
Qy 1341 SerProH1eH1eYsG1YGlnH1eH1e1aArgG1YSer1YleThrGlnG1Y1ePro 1360
Db 4022 AGCCCGACACCTCAAAAGAGCAGCAGCATCCGCGGTCCATCACAAGGATCTCT 4081
Qy 1361 ArgSerTYrValG1YAlaGlnG1YAspTYrLeuArgArgG1YAlaYsLeuLeuLYsArg 1380
Db 4082 CGGTCTTACGTGAGGACAGAGAGCATCTGCTCGGAGGCGCAAGCTCTTAAAGCGG 4141
Qy 1381 G1YG1YThrProProProProProProSerArgAspLeuThrG1YAlaTYrLYsThGln 1400
Db 4142 GAGGACAGCTCGCCCGCCACCGCTCAGGAGACTGACGAGGCTTCAAGACCCAG 4201
Qy 1401 AlaLeuG1YProLeuYsLeuYsProAlaH1eG1YLeuValAlaThraLYsG1Y 1420
Db 4202 GCTCTGGCGCCCTGAAGCTGAAGCGGCCCAAGAGGCTGTGTGCGCAGGTGAAGAG 4261
Qy 1421 AlaG1YArgSer1YleH1eG1Y1eProArgG1YLeuArgH1eThrProG1YLeuPro 1440
Db 4262 GCGGCGCGCTCATCATGATGATCCCGGAGAGACTGTGGGACAGCGCCGAGCTGCC 4321
Qy 1441 LeuAlaProArgProLeuYsG1YSer1YleThrGlnG1YThrProLeuYsTYrAsp 1460
Db 4322 CTGGCGCCGGGCGCTCAAGAGGCTCCATCAGCAGGAGCCCGGCTCAAGTACGAC 4381
Qy 1461 ThrG1YAlaSerThrThrG1YSerLYsLYsH1aAspValArgSerLeu1YleG1YSerPro 1480
Db 4382 ACCGGGCGTCCACCATGTGCTCAAAAGAGAGAGTACGCTCCCTCATCGGCAAGCCCC 4441
Qy 1481 G1YArgThrPheProProValH1eProLeuArpValMetAlaArpAlaLeuG1Y 1500
Db 4442 GGCCTGAGCTTCCACCCGTGCACCGCTGATGTATGGCGAGCCCGGCACTGGAA 4501
Qy 1501 ArgAlaCYeTYrGlnG1YSerLeuYsSerArgProG1YThra1aSerSerSerG1YG1Y 1520
Db 4502 CGTGTGTCTACAGAGAGCTGAAAGCGCGGAGAGCGCAGAGACTGGGGGCG 4561
Qy 1521 Ser1YleAlaArgG1YAlaProVal1YleValProG1YLeuG1YLYsProArgGlnSerPro 1540
Db 4562 TCATATTGCGCGCGCGCCCGGCTCATTTGTGCTGAGACTGGTAAAGCGCGAGAGCCCC 4621
Qy 1541 LeuThTYrGlnArpH1eG1YAlaProPheAlaG1YH1eLeuProArgG1YSerProVal 1560
Db 4622 CTGACCTTAGAGAGCAGCGGAGCACTTTGGCGGACCTCCACAGAGGTTGGCCGTG 4681
Qy 1561 ThrMetArgG1YProThrProArgLeuGlnG1YSerLeuSerSerSerLYsAlaSer 1580
Db 4682 ACCATGGGAGAGCCAGCGCGGCTGCAAGAGGAGGCTTTCGTCAGCAAGGATCC 4741
Qy 1581 GlnAspArgLYsLeuThSerThrProArgG1YAlaLYsSerProH1eSerThraVal 1600
Db 4742 CAGAGCCGAAGCTAGCTGAGAGCTGTGATGATGCCAAGTCCCGCAGCAGCAGCTG 4801
Qy 1601 ProG1YH1eH1eProH1ePro1YleSerProTYrGlnH1eLeuLeuArgG1YAlaSerG1Y 1620
Db 4802 CCGAGACACACCAACCCCATCTGCGCTTATGAGACCTGTGGGGCGTGAAGTGGC 4861

1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
4862 GTGGACCTGTATTCGACGCACATCCCTCGGCTTCGACCCACCTCCATACCCCGGCGC 4921
1641 IleProLeuAspAlaAlaAlaIleTyrTyrLeuProArgHisIleuAlaProAsnProThr 1660
4922 ATCCCTGTGAGCGAGCGCTGCTACTACTGCTCCCGGACACCTGGCCCCCAACCCGAC 4981
1661 TyrProHisIleuTyrProProTyrLeuIleArgGlyTyrTyrProAspThrAlaIleuGlu 1680
4982 TACCCGCACTGTATCCACCTCACTCACTCGGCGGTACCCGACACGGCGGGCTGGAG 5041
1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnIleMetHisIleAsnThrAla 1700
5042 AACCGGAGACCATCATCATGACTATGACTACCTCGGACAGATGACACCAACACGGGCGC 5101
1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu 1720
5102 ACCGCATATGGCCAGCGACTGATATGATGAGGGGCGCTCGCGCCGCGAGATCTCGCTG 5161
1721 AlaLeuAsnTyrAlaAlaIleTyrProArgGlyIleIleAspLeuSerGlnAlaProHisIleu 1740
5162 GCACTCACTAGCTGCGGGGTCCCGAGGCAATCATGACCTGTCCCAAGTGGCCACACTG 5221
1741 ProValIleuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
5222 CTTGTGCTGTCGCCCCCGGACACAGGCAACCCGACCGCATGAGCGGCTGCTGCTAC 5281
1761 LeuProThrAlaProGlnProPheSerSerArgHisIleSerSerSerProLeuSerProGly 1780
5282 CTCCCGACCGCGCCCGGACCTTCAGACACCGGCAAGAGCTCCCACTCTCCCGAGCA 5341
1781 GlyProThrHisIleuThrTyrSerProThrThrThrSerSerSerGluArgGluArgAspArg 1800
5342 GGTCCAAACACTTGCACAAACCAACACACGCTCTCGTCCGAGGGGAGGAGGACCG 5401
1801 AspArgGluArgAspArgAspArgGluArgGluArgSerIleuThrSerThrThrThr 1820
5402 GATCCAGAGACCGGACCGGATCGGAGCGGGAATACTCATCTCCATCGTCCACACAGAG 5461
1821 ValGluHisAlaProIleTyrProGlyThrGluGlnSerSerGlySerSerGlySer 1840
5462 GTGGAGCAGCAGCAGCCTCTGAGACCTGTGTACAGACAGACAGCGCGCAGCGCAGC 5521
1841 SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
5522 ACGCGCGGGGTGGGGGAGCAGCAGCAGCGCGCTCCCATCTCCATGCGCCACAGCAC 5581
1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
5582 TCGCCCATCTCCCTCGGACCCAGCATGCTCTCACAGAGACCCAGTGTGCTTCACAA 5641
1881 ThrGlyMetLeuGlyIleIleThrAlaValGluProSerIleProThrValLeuArgSer 1900
5642 AAGAGCATGAAGGGATCATCAACGCTGTGAGACCCAGCAAGCCCAAGCTCTGAGAGTCC 5701
1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
5702 ACCCTCACTCTCCACCGCTTCGCCAGAGTGCACATCTCCACCTCGACCCACCTCCCA 5761
1921 LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValIleuLeuProLys 1940
5762 CTGGGGGGGACCTTCATGGGGGTCTACCCCTCATGAGGCGCGCTTGGCTGCCCAAG 5821
1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
5822 GAGGCGCCCGGGGTCCCGCGCAGAGCGGCGCCGAGCAGACACCGGCACTGCTTCTC 5881
1961 AlaIleProProAlaArgSerGlyLeuGluProAlaSerSerProSerIleArgGlySerGlu 1980
5882 GCCAACCCCGCAACCGCGCTCGGGCTGGAGCCCGCTCTCCCGCAGCAAGGGCTCGAG 5941

1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys 2000
5942 CCGCGCCCTGTAGCTCTCTGTCTGTGACCGCACCATGATGCCGCAACCCCTCGAAG 6001
2001 AsnLeuAlaProHisIleAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
6002 AACCTTGACCTTACACAGCAGCCAGCGGAGCCCGCGGCGCCACTGCTCGGCTCGGAC 6061
2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSer 2040
6062 CCGCACCGGGAABAAGCTCAAGTAACCCCTTTCATCTCAGGAACCTGGAACCTCGCTCT 6121
2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
6122 CTGGGTATCAACGAGCAGAGCTACACCCCGAAGGGGTGGAGCCCGTCAAGCTGTGAGC 6181
2061 SerProSerLeuThrHisAspLysGlyLeuProLysProLysHisIleuGluLeuAspLysSer 2080
6182 TCACCCAGTGTGACCAACAGAGGGGCTCCCAAGCACTTGAAAGAGCTTCAGCAAGAGC 6241
2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValIleuLeuGlyGlyAla 2100
6242 CACTCGAGGGGGAGCTGGGCGCAAGAGCAGGCGCCGTGAGCTTGGCGGGAGGCGC 6301
2101 AlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
6302 GCCCACTTCCACACTCGGCGCGCTGCTGTAGAGCCAGCCCTGTCAGCGCGCTGCTC 6361
2121 GlnThrAlaProGlyValIleArgHisIleGlnArgValIleThrLeuAlaGlnHisIleSer 2140
6362 CAGACCGCCCGAGGGGTCAAAAGTACACAGCGGGGTGTCACTTGCCACAGCACAATCAT 6421
2141 GluValIleThrGlnAspTyrThrArgHisIleProGlnGlnLeuSerAlaProLeuPro 2160
6422 GAGGTATACACAGAGACTACACCGGCAACCCAGCAGCACTCAGCGCACCCCTGCCC 6481
2161 AlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProPro 2180
6482 GCCCCCTCTACTCTCTCTCTGCGGCGACCTCCCGTCTGTGACTTCGCGCGCCACCC 6541
2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
6542 AGTGAACCTTACTCTCCGCGCCGAGACATGGTGGCCCGGTGCTCTCCCGCACAGC 6601
2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220
6602 GAAGGGGAGCAAGGCTCTCAGAGCCAAAGACGTGCGTCTTGGGTGTGGTGAAGAC 6661
2221 GlyIleGluProValSerProProGlyGlyMetThrGluProGlyHisSerArgSerAla 2240
6662 GGATATTGAACCTGTGTCCCAACCGGAGGAGCATGACGAGGCCAGGCACTCCCGGAGTGC 6721
2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys 2260
6722 GTGTACCGCTGTATCCGGATGGGAAACAGACGAGCCACAGCATGGGGCTCCAA 6781
2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
6782 TCTCAGGAGCAACACAGCAGCGGCGCAGCTTCTTACAGAACTGACGAGAGCACTCC 6841
2281 AlaMetValLysSerLysIleGlnGluIleAsnLysLysLeuAsnThrHisAsnAlaGln 2300
6842 GCCATGTCAAGTCCCAAGAGAGCAAGATCAACAGAACTTAAACCCCAACACCGGAAT 6901
2301 GluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr 2320
6902 GAGCTGGAATACATATACGCGAGCTGGAGAGGAGATTTCAATATGCCCCGCAACACC 6961
2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisIleAsnThrAsnMet 2340
6962 GGAACAGGCTTATGACTATATGAGCGCAGCGGTGCAAGAAATCAGCAACCAACATG 7021
2341 GlyLeuGluAlaIleIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSer 2360

|||||
Db 7022 GGGCTGAGGCGCATATTTAGAAAGCACTCATGGGTAAATATGACCAAGTGGGAAGATCC 7081
Qy 2261 ProProLeuSerAlaAanAlaPheAsnProLeuAenAlaSerAlaSerLeuProAlaAla 2380
Db 7082 CCGCGGCTCAGCGCAATGCTTTAACTCTGATGACATGTCAGAGCTTCCGCTGCT 7141
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
Db 7142 ATGCCATTAACCGCTGCTGAGAGAGATGACCAACACTCACTCCGACAGTGGCGGC 7201
Qy 2401 GlyValAlaValSerGlyArgProSerSerArglyValAlaSerProAlaProGly 2420
Db 7202 GGGAAAGCCAAAGCTCTGTGGCAGACCCAGCAAGCAAGCAAGTCCCGCCCGGC 7261
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlyGlyAspCysAsn 2440
Db 7262 CTGGCATCTGGGAGCCGCGCACCTCTGTCTCTCAGTGCATCTGGAGGAGACTGGCAAC 7321
Qy 2441 ArgArgThrProLeuThrAsnArgValTropGluAspArgProSerSerAlaGlySerThr 2460
Db 7322 CGCGGAGCGCGCTCACCACCGCGTGTGGAGAGACAGCCCTGCTCCGCAAGTTCCACG 7381
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7382 CCAATCCCTTACAAACCCCTGATCATGCGGCTGCAGCGGGGTGCATGGCTTCCCAACC 7441
Qy 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaAlaTyrAspGlu 2500
Db 7442 CACACCGGCTCTCCCGGCGAGCGGGCCCTCTGCTGCGCCCAACGCGCTGGAGACGAG 7501
Qy 2501 GluProAlaProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 7502 GAACCCAGCCACTGCTCTGCTCCGATACGAGACACTCTCCGACAGCGAG 7552

RESULT 3
US-11-052-344-1
; Sequence 1, Application US/11052344
; Publication No. US20050191674A1
; GENERAL INFORMATION:
; APPLICANT: Cognate, Lionel
; TITLE OF INVENTION: Method Of Prognosis Of Metastasis By Detection Of FRA12E Fragile
; FILE REFERENCE: 03551.0181
; CURRENT APPLICATION NUMBER: US/11/052,344
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/542,538
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 1
; SEQ ID NO 1
; LENGTH: 8561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SMRT gene which includes the open reading frame for SMRT protein.
; OTHER INFORMATION: The open reading frame starts at nucleotide 2 and ends at nucleot
US-11-052-344-1

Alignment Scores:
Pred. No.: 0 Length: 8561
Score: 13215.00 Matches: 2517
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 26 Gaps: 0

US-09-522-753-5 (1-2517) x US-11-052-344-1 (1-8561)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrArgLysProArgTyrPro 20
Db 2 ATGTGGGCTCCACACAGCTTTGTGGCACAGCGTGGGCGGACCTGAGCCCGCTACCG 61
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40

|||||
Db 62 CCCACAGCCTTCTCTACCCAGTGCAGATGCCCGGACGACACAGAGCTGGGCTCTG 121
Qy 41 GlnTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerTlleGln 60
Db 122 GAGTACAGACACCACTCCCGGACTATGCTCCCACTGTGGCGGCTCATATCCAG 181
Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 182 CCCAGGCGCGAGGCCCTCTCTGCTGTGATTCAGATCCCGGGAATGAACGATCCAG 241
Qy 81 GlnLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlySerGluMet 100
Db 242 GAGCTCACCTCGCGGACAGAGTCCCATCACTGCCCGGAGCTGGGGAATCAGAGATG 301
Qy 101 GluPheIleGlySerTyrValProArgLeuGluLeuLeuProAlaProLeuLeuArgPro 120
Db 302 GAGTTCAATTAAACCAAGCCCTCGGCTAGAGCTGTGCTGACCCCTGCTCGACCG 361
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 362 TCACCTCTGCTGCGCACGGGCGAGCTGCGGATCTTAAGACTCACCAAGACCTGATGC 421
Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160
Db 422 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGGACCCCGGACACTGACCTGAGCTG 481
Qy 161 GluLeuValProProAlaGlySerTyrGlyGluLeuIleGlnAsnMetAspArgValAsp 180
Db 482 GAGCTGTGTGCGCGCACCGCTGTCCMAAGAGAGCTGTATCCAGAACATGACCGGTGGAC 541
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerTyrLeuTyrLysGlnGlnGlnLeu 200
Db 542 CGAGAGTACACCATGTATGAGACAGACATCTTAAGTGAAGAAAGACACACAGCTG 601
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
Db 602 GAGGAGAGGCTGCGCAAGCCCGGACCTGAGAAAGCCCGTGCACCGCGCCATCGAG 661
Qy 221 SerTyrHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValGlnAla 240
Db 662 TCAGAGCACCGCAGCCGTGTGACATCATACGACAGAAACCGGAAAGGCTGAAAGCT 721
Qy 241 AlHisArgTlleLeuGluGlyLeuGlyProGluValGluLeuProLeuTyrTyrAsnGluPro 260
Db 722 GCACATGCAATTTCTGAAGGCTGGGCCCCAGGTGAGCTGCCCTGTACACCAAGCCC 781
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
Db 782 TCCGACACCCGGCAGATCATGAGAACATCAAAATAAACCAAGCGGATGCCGAAGACCTA 841
Qy 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnIleTyrLysGlnLysPheCysGln 300
Db 842 ATCTTGACTTCAAGAGAGAAATCAAGCTCGGAAACAAAGGAAGCAAGAGTTCTGCAG 901
Qy 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluTyrGlnIleGluAsnAspPro 320
Db 902 CGCTATGACCAAGCTCATGAGGCTTGGAAAATAAGGTGAAGGACATCGAAACACCCG 961
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db 962 CGCGCGGGGCCAAAGAGAGACAAAGTGGCGGAGTACGAAAAAGCAAGTCCCTGAGATC 1021
Qy 341 ArgLysGlnArgGluLeuGlnGlnLysMetGlnSerArgValGlyGlnArgLysGly 360
Db 1022 CGCAACAGCGCAGCTGCAGAGCCATGCAAGACAGAGGTGGGCCGCGGCGCAGTGGG 1081
Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluTyrIleIleAspGlyLeuSer 380
Db 1082 CTGTTCATGTGCGGCCCGCCGACGACGACAGAGTGTCAAGATCATGTGCTCTCA 1141
Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400

Dh 1142 GAGCAGAGAACCTGGAGAGCAGATGCCGACGCTGGATCCCGCCATCGTGTAC 1201
Qy AASPAlaAspGlnGlnArgIleIysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
Dh 401 |||||
Dh 1202 GAGCGTGAACAGAGGACATCAAGTTATCAACATGAACGGCTTATGGCCGACCCATG 1261
Qy 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnGlyIleIysPhe 440
Dh 1262 AAGGTGTACAAGACCGCCAGGTCAATGACATGTGGATGAGCAGAGGAAGAGACTTTC 1321
Qy 441 ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
Dh 1322 CGGGAGAAAGTTCATGACGATCCCAAGAACTTTGGCTGATCCCATTCCTCGAGAGG 1381
Qy 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnAsnTyrLys 480
Dh 1382 AAGACGTGGCTGAGTGCCTCTCTTACTTACTGACTAGAGAAATGAGAACTTAAAG 1441
Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Dh 1442 AGCTGTGTGAGAGAGCTATCGGCCCGCGCAAGAGCCAGCAGCAACAGCAGCAG 1501
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Dh 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
Qy 521 AspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 540
Dh 1562 GATGAG 1621
Qy 541 AspLysGlnLysPheLeuLysGlnLysTyrAspAspThrSerGlnGlnAspAspGln 560
Dh 1622 GACAGAGAGAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
Qy 561 LysGlnLysValAlaLysLysGlnLysGlnLysThrAlaAsnSerGlnGlnLysGln 580
Dh 1682 AAGAGAGCTGTGCTTCCAAAGGCCCGCAAACTGCCAAACAGCCAGAGAAACCCAAAGGC 1741
Qy 581 ArgIleThrArgSerMetAlaAsnGlnLysAlaAsnSerGlnGlnLysIleThrProGln 600
Dh 1742 CGCATCAACCGCTCAATGGCTTAATGAGGCCAACAGAGAGAGAGAGAGAGAGAG 1801
Qy 601 SerAlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 620
Dh 1802 AGGCGCGAGCTGCGCTTCATGAGAGCTGAATGAGAGTTCGCTGAGCAGAGAAAGAAATG 1861
Qy 621 GlnThrAlaLysLysGlnLysLysGlnLysGlnLysLysLysLysLysLysLysLys 640
Dh 1862 GAAACAGCCCAAGAAAGGTCTCTGGAACAGCGCCGCAACTGTGCGGCATCGCCGATG 1921
Qy 641 ValGlnSerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
Dh 1922 GTGGGCTCCAGACTGTGTGCGAGTGAAGACTTCACTTCAACTCAAGAAAGAGGACG 1981
Qy 661 AsnLeuAspGlnLysLeuGlnGlnLysLysLysLysLysLysLysLysLysLysLys 680
Dh 1982 AACCTGATGAGATCTTGCAGCAGCAGCAACAGCTGAATGAGAGAGAGAGAGAGAG 2041
Qy 681 ArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 700
Dh 2042 AGGAAAGAAAGAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2101
Qy 701 GlnAspGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 720
Dh 2102 GAGGATGAGAGAGATGAGAGCGCTGCGCGCTGAGCGGAAATGAGAGAGAGAGAGAG 2161
Qy 721 AlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 740
Dh 2162 GCTGAAGCCTTATCATGCTCTGGGAATGAGAGTCCCAAGAGGAGATCAAGGCGCCAGCC 2221
Qy 741 ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisGlnGlnLysLys 760
Dh 2222 ACTGTCAACAAAGCTCAGACCGAGAGATCCCTCTCTCTCACTGAGGCGCGCCAG 2281

Qy 761 AspThrGlnLysLysGlnLysProLysProProAlaThrLeuGlnValAspGlnLysProPro 780
Dh 2282 GACACAGGCGAGATGGCGCCAAAGCCCGCAGCACTCTGGCGCGCAGCGGCAACCCCA 2341
Qy 781 GlyProProThrProProAsnArgThrSerArgAlaProIleGlnLysProProAlaSer 800
Dh 2342 GCGCCCAACCAACCCCAACAGAGAGATCCGGGCCCCCATTTGAGCCCAACCCCGGCTCT 2401
Qy 801 GlnAlaThrGlnValaProThrProProProAlaProProSerProSerAlaProProPro 820
Dh 2402 GAAGCCACCGAGAGCCCTTACGCCCCACAGACCCCATTCGCTCTGACCTCTCTCT 2461
Qy 821 ValLysLysGlnLysLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 840
Dh 2462 GTGTCTCCCAAG 2521
Qy 841 GlnGlnGlnLysProProAlaAlaGlnGlnLysLysLysLysLysLysLysLysLys 860
Dh 2522 GAGAGCAG 2581
Qy 861 ProValLysSerGlnCysThrGlnGlnLysGlnLysGlnLysGlnLysGlnLysGln 880
Dh 2582 CCGTCAAG 2641
Qy 881 GlnAlaLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 900
Dh 2642 GAGCGCGCTGAGGCGCAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2701
Qy 901 GlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 920
Dh 2702 GGCAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2761
Qy 921 CysSerAlaAspGlnLysLysGlnLysGlnLysGlnLysLysLysLysLysLysLys 940
Dh 2762 TGCACTGACAGAGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2821
Qy 941 ArgProSerLeuLeuThrProThrGlnLysAspProArgAlaAsnLysSerProGlnLys 960
Dh 2822 AGGCCAGCCTCTCAACCCGACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2881
Qy 961 LeuAspLeuLysGlnLysLysGlnLysGlnLysLysLysLysLysLysLysLysLys 980
Dh 2882 CTGGACCTTAAG 2941
Qy 981 ValHisGlnProProArgGlnAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Dh 2942 GTTCATGAG 3001
Qy 1001 ProProGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1020
Dh 3002 CCACCGCAAAACCTGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1021 GlnLysSerArgSerProAlaProProAlaAspLysGlnLysLysLysLysLysLys 1040
Dh 3062 GGCMAAG 3121
Qy 1041 LysLeuProGlnLysAspProProCysTrpThrSerGlnLysProPheProValProProArg 1060
Dh 3122 AAGCTCTGAGGAG 3181
Qy 1061 GlnValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1080
Dh 3182 GAGGTATCAAG 3241
Qy 1081 GlnHisProLeuProLeuGlnLysLysLysLysLysLysLysLysLysLysLysLys 1100
Dh 3242 GGTCAACCACTGAG 3301
Qy 1101 ThrIleSerAsnProProProLeuLysSerSerAlaLysHisProSerValLeuGlnLys 1120
Dh 3302 ACCATCTCAACCGAGCT 3361

QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlnHis 1140
DB 3362 CAATATAGTGCATCTCCCAAGAAATGTCGGTCCAGCTCCACGTCCTACTCAGACAT 3421
QY 1141 AlAlaValAlaProValGlyProValThrmecGlyLeuProLeuProMetLeuProLys 1160
DB 3422 GCCAAGGCCCCGGGGCCCTGTACATGGGGCTGCCCTGCCATGAGACCCCAAAAG 3481
QY 1161 LeuAlaProPheSerGlyValLysGlnGlyLeuLeuSerProArgGlyGlnAlaGlyPro 1180
DB 3482 CTGGACACCTTCACGGAGTGAAGCAAGAGAGCTGTCCCAAGGGGCGAGCTGGGCCA 3541
QY 1181 ProGlnSerLeuGlyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeu 1200
DB 3542 CCGGAGAGCTGGGGGGTGGCCACAGGCCCAAGAGGCGTCCGTGTGAGAGGAGCAGCTCTG 3601
QY 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
DB 3602 GGCTCAGTTCGGGGCGGAGCATTCACCAAGGCAATCCAGCACAGGAGTGCCTCGGAC 3661
QY 1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValIleuTyrLys 1240
DB 3662 AGCGCCATCAATACCGGGCTCATACCAAGGCAAGCGCATGTCGTCTGTACAG 3721
QY 1241 GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu 1260
DB 3722 GGCAACATCAAGAGATCATCGGAGAGACAGCCGAGTGTGAGCCGGGGCGGAG 3781
QY 1261 AspSerLeuProLysGlyHisValIleTyrGlyGlyLysLysGlyHisValLeuSerTyr 1280
DB 3782 GACAGCTGCGCCAAAGGCGCATGTATCAAGAGGCAAGAGGCGCATGTCCTAT 3841
QY 1281 GlnGlyGlyMetSerValThrGlnCysSerLeuGlnAspGlyArgSerSerSerGlyPro 1300
DB 3842 GAGAGTGGCATGTCTGTGACCAAGTGTCCAAAGAGACGCGCAAGACACTCAGAGCC 3901
QY 1301 ProHisGlyThrAlaAlaProLysArgThrTyrAspMetMetGlyArgValGlyArg 1320
DB 3902 CCCCATGAGACGGCGCGCCCAAGCGCATGATGACATGAGAGGGCGCGTGGGAGA 3961
QY 1321 AlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIleProProGlnArgHis 1340
DB 3962 GCCATCTCTCCACCCAGCATCGAAGTCTCATGGGCGGTCCATCCCGCGAGCGCAC 4021
QY 1341 SerProHisIleLeuLysGlnHisIleHisIleArgGlySerIleThrGlnGlyIlePro 1360
DB 4022 AGCCCCACCACTTCAAGAGACAGCACACATCGCGAGTCTACACACAGAGATCCT 4081
QY 1361 ArgSerTyrValGlnAlaGlnGluAspTyrLeuArgArgGlyAlaLysLeuLeuLysArg 1380
DB 4082 CGGTCTTACGTGAGGACAGAGGACTTACTGTGCGGAGGCGCAAGCTCTTAAAGCGG 4141
QY 1381 GlnGlyThrProProProProProProProProProProProProProProProPro 1400
DB 4142 GAGGAGCAGCGCTCGCGCCCAAGCGCTTCAAGGAGCTGACCGAGGCTTCAAGAGCGCAG 4201
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlyGlyLeuValAlaThrValLysGln 1420
DB 4202 GCCCTGGGCCCCCTGAAGCTGAAGCGGAGCCCATGAGGGCTGTGGGCGACAGGAGAG 4261
QY 1421 AlaGlyArgSerIleHisGlnIleProArgGlyGlyLeuValArgHisThrProGlnLeuPro 1440
DB 4262 GCGGGCGCGCTCATCATGAGATCCGCGGAGAGCTGGGACACGCGCGAGCTGCC 4321
QY 1441 LeuAlaProArgProLeuLysGlyGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
DB 4322 CTGGGCCCGCGCGCTCAAGAGGAGCTCATACGAGGAGCAAGCGCTCAAGTACGAC 4381
QY 1461 ThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerPro 1480
DB 4382 ACCGGGCGCTCCACCACTGGCTCCAAAGACAGAGTACGCTTCCCTCATGGGAGGCC 4441
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGln 1500

DB 4442 GCGCGAGCTTCCCAAGCGTGCAGACCCGCTGATGTATGAGCCAGCGCCGGGCACTGGAA 4501
QY 1501 ArgAlaIleTyrGlyGlnGlySerLeuLysSerArgProGlyThrAlaSerSerSerGly 1520
DB 4502 GTGTCTGTACAGAGGAGACTGTGAAGAGCGGCGAGAGACCGCACACTTCGGGGGCG 4561
QY 1521 SerIleAlaArgGlyAlaProValIleValProGlnLeuGlyLysProArgGlnSerPro 1540
DB 4562 TCCATTGCGGGCGCGCGCGCTGATTGTGCTGAGCTGGGTAAAGCCGGCGAGAGCTCC 4621
QY 1541 LeuThrTyrGlnAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
DB 4622 CTGACCTATGAGAGCACAGGGGACACTTGGCGGACACTCCACAGAGTTCGCGCGTG 4681
QY 1561 ThrMetArgGlnProThrProArgLeuGlnGlySerLeuSerSerSerLysValSer 1580
DB 4682 ACCATGGGAGGCCACAGCGCGCTGCGAGAGGAGCGCTTGTCTCACCAAGGCAATCC 4741
QY 1581 GlnAspArgLysLeuThrSerThrProArgGlyIleAlaLysSerProHisSerThrVal 1600
DB 4742 CAGGACGAAAGCTGAGCTGAGCGCTCTGTGAATGCGCAAGTCCCGGACAGACCGTG 4801
QY 1601 ProGlnHisIleProHisProIleSerProTyrGlnHisLeuLeuArgGlyLysGly 1620
DB 4802 CCGAGACCAACCAACCCCATCTGCGCTTATGAGACCTGCTGGGGCGTGAAGTGC 4861
QY 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
DB 4862 GTGACCTGTATGAGCGCCATCCCTGCGCTTGCACCCACCTTCATACCCCGGCG 4921
QY 1641 IleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisIleuAlaProAspProThr 1660
DB 4922 ATCCCTTGAGCCAGCGCGTGTCTTACTCTCCCGAGACCTTGCGCCCAACCCAC 4981
QY 1661 TyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGln 1680
DB 4982 TACCGGACCTGATACCAACCTTACTCTATCGGGGCTACCCGACAGCGGGCGGTGAG 5041
QY 1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisIleAsnThrAla 1700
DB 5042 AACCGGACAGCAATCATGATGATCATCACTTCCAGAGATGACCAACCAACCGGCC 5101
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSerLeu 1720
DB 5102 ACCGCGATGGCCAGAGAGCTGATATGCTGAGGGGCTCTCGCGCGGAGTCTCGCTG 5161
QY 1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
DB 5162 GCACTCAACTACGCTCGGGTCCCGAGGACATGACCTGTCCCAAGTGCACACCTG 5221
QY 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
DB 5222 CTGTGCTGTGCGCCCGGACACAGGACCCCAAGCGCATGAGACCGCTTGGCTTAC 5281
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
DB 5282 CTCCCAAGCGGCCAGAGCCCTTACAGAGCGGACAGAGCTTCCCACTTCCCAAGGA 5341
QY 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGlyLysArgLysArg 1800
DB 5342 GGTCAACACACTTGAACAAACCAACACACAGTCTCGTCCGAGCGGAGCGAGACCGG 5401
QY 1801 AspArgGlnArgAspArgAspArgGlyLysGlyLysSerIleLeuThrSerThrThrThr 1820
DB 5402 GATCGAGACCGGAGCGGAGTGGGAGCGGAAATCTCATCTTCAAGTCCACAGACG 5461
QY 1821 ValGlnHisAlaProIleThrParProGlyThrGlnGlnSerSerGlySerSerGlySer 1840
DB 5462 GTGGAGCAGCAACCACTTGGAGACTGTATCAGAGCAAGACAGCGGCGAGCGGACG 5521
QY 1841 SerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860

5522 AGCGGCGGGGTGGGGGACAGACCGCCCGCTCCACTCCCATGCCACGACGAC 5581
1861 SerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsn 1880
5582 TCGCCATCTCCCTCGAGCCAGATGCCCTCCACAGAGACCGATGCTTCCACAC 5641
1881 ThrGlyMetIleGlyIleIleThrAlaValGluProSerIleProThrValLeuArgSer 1900
5642 ACAGGATAGAGGGATCATCACCGCTGTGAGACCCAGACGACCCAGCGCTCGAGGTCC 5701
1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
5702 ACCCTCACCCTCTCCACCGCTTGCCTGCACAGCTGCACATCTCCACCTCCACCTCCCA 5761
1921 LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys 1940
5762 CTGGGGGGGACCTCGATGGGGGTCTACCTCATCGAGGCGGTCTTGCTGCCCAAG 5821
1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
5822 GAGGCCCCCGGGGTGCGCGGCGACAGCGGGCCCGACAGACACCGGCGCATGCTTCTC 5881
1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerIleGlySerGlu 1980
5882 GCCAAGCCCCCAGCCCGCTCCGGGCTGAGCCCGCTCTCTCCCGACAGAGGGCTCGAG 5941
1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys 2000
5942 CCCCCGCGCTAGTGCCTCTCTCTGTGGCGACGACCATGCGCCGACCTCGGGGAAAG 6001
2001 AsnLeuAlaProHisIleAspSerProAspProProAlaProProAlaSerAlaSerAsp 2020
6002 AACCTGCACCTTCACACGCGACCGGAGCCGCGCGCGCACCTGCTCGGCTCGGAGC 6061
2021 ProHisArgGlyLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSer 2040
6062 CCGCACCGGGGAAAGACTCAAGTAAACCTTTTCATCCAGGAACTGGAACTCCCTTCT 6121
2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
6122 CTGGGTTACCGAGCGACGACTACAGCCCCGAGAGGGGTGAGGCCCTGACCTGTGAGC 6181
2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLysLeuAspLysSer 2080
6182 TCAACCAAGTGTGACCAACGACGAGGGGCTCCCAAGACCTGGAAGAGCTCGACAGAGC 6241
2081 HisLeuGlnGlyGlyLeuArgProLysGlnProGlyProValLysLeuGlyGlyValAla 2100
6242 CACTGTGAGGGGAGCTGGCGCCAGACGACGCGCCGCTGAAGCTTGGCGGGAGGCC 6301
2101 AlaHisLeuProHisLeuArgProLeuArgProGluSerGlnProSerSerSerProLeuLeu 2120
6302 GCCCACCCTCCACACTTGGCGCGCTGCTGAGAGCCAGCCCTCTGTCAGCCCGCTGCTC 6361
2121 GlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
6362 CAGACCGGCCCAAGGGGTCAAGGTCAACAGCGGTGTCACTTGGCCCGACATCAAGT 6421
2141 GluValIleThrGlnAspTyrThrArgHisAspProGlnLeuSerAlaProLeuPro 2160
6422 GAGGTATATCACAGAGCTACACCGGACACCCACAGAGTTCAGCGGACCCCTGCGCC 6481
2161 AlaProLeuTyrSerPheProGlyValAspCysProValLeuAspLeuArgArgProPro 2180
6482 GCCCCCCCTCTACTCTTCTTCTGGGGCCAGGTGCGCTCTGGAAGCTTCCGCGCCACCC 6541
2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
6542 AGTGACTGTACTCCCGCGCCCGGACCATGTGTGCCCGCGCTGCTCGCCACAGC 6601
2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyValAsp 2220
6602 GAGGGGGGCAAGGTCTCCAGAGCCAAACAGACGTGGTCTTGGGTGTGTGAGGAC 6661

2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
6662 GGTAATTGAACCTGTGTCCCAACCGGAGGGCATGACGAGACCGAGGACATCCCGGAGTGCT 6721
2241 ValTyrProLeuLeuTyrArgAspGlyGlyGlnThrGluProSerArgMetLysLys 2260
6722 GTGTACCGCGCTGTGTACCGGATGGGAAACAGACGGAGCCGACAGATGGGCTTCCAAAG 6781
2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
6782 TCTCAGGCAACACACGACGCCGCCAGCTTCTTACGACAGACTACCGAGGCAACTCC 6841
2281 AlaMetValLysSerLysLysGlnLysLeuAsnLysLeuAsnThrHisAsnArgAsn 2300
6842 GCCATGTCTAAGTCCAAAGAGCAAGAGATCAACAGAGCTGAACACCAACCGGAAT 6901
2301 GluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr 2320
6902 GAGCTGTAAATACAAATATCAGCGCTGGAGCGAGATCTTCAATATGCTCCGCTATCAC 6961
2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet 2340
6962 GGAACAGGCTTATATGACTATAGAAAGCCAGGCGGTGCAAGACATGCCAGCAACAATG 7021
2341 GlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSer 2360
7022 GGGCTGAGGCGCATATATGAAAGCACATCAGGGTAAATATAGACACAGTGGAGAGTCC 7081
2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
7082 CCGCGCTCAGGCGCAATGCTTTTAACTCTGAAATGCCAGAGCTGCGCGCTGCT 7141
2381 MetProIleThrAlaAlaAspGlyArgSerAspHisAsnLeuThrSerProGlyGly 2400
7142 ATGCCATTAACCGCTGCTGACGAGAGTACACACTACCTCCGCAAGTGGCGGC 7201
2401 GlyValAlaLysValSerGlyArgProSerSerArgLysValLysSerProAlaProGly 2420
7202 GGAAGAGCCAAAGTCTCTGCGACAGCCAGCACCGGAAAGCCAAAGTCCCGCGCGCGC 7261
2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyLysAspAsn 2440
7262 CTGGGATCTGGGAGACCGGACCTCTGTCTCTCACTGACACTCGAGGAGACTGCAAC 7321
2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
7322 CCGCGGAGCGCGCTACCAACCGCGTGGAGAGACGSCCTCGTCCGACAGTTCCACG 7381
2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
7382 CCATTCCTTACAAACCCCTGATCATGCGGCTGACAGGCGGGTGTCTATGCTTCCACACC 7441
2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaIleTyrAspGlu 2500
7442 CACCGGCGCTCCCGCGGAGCGGGCGCTGCTGCTGCGCCACCAACCTCGGAGCGAG 7501
2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
7502 GAGCCCAAGCCACTGCTCTGCTCGCATGACGAGACCTTCCGACAGCGAG 7552

RESULT 4
US-10-087-192-653
; Sequence 653, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ. ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 653
LENGTH: 8667
TYPE: DNA
ORGANISM: Homo sapiens
US-10-087-192-653

Alignment Scores:
Pred. No.: 0 Length: 8667
Score: 13145.00 Matches: 2509
Percent Similarity: 99.68% Conservative: 1
Best Local Similarity: 99.64% Mismatches: 7
Query Match: 99.47% Indels: 2
DB: 14 Gaps: 0

US-09-522-753-5 (1-2517) x US-10-087-192-653 (1-8667)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpAlaGlnAlaThrGluProAlaGlyTyrPro 20
DB 116 ATGTGGGATCCACACAGCCCTGTGGCACAGCTGGAGGGCCACTGAGCCCGCTACCCG 175

QY 21 ProHisSerLeuSerTyrProValGlnLeuAlaArgThrHisThrAspValGlyLeuLeu 40
DB 176 CCCACAGCCTTTCTTCCACAGTGCAGATGCCCCGACGACACAGAGCGTGGGCTCTCG 235

QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 236 GAGTACAGACACCACTCCCGCACTAGCTCCACCTGTGCGGCTCCATCATCAG 235

QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 296 CCCACAGCGGGAGGCCCTCTGCTGTCTGAGTTCCAGCCCGGAAATGAACGGTCCAG 355

QY 81 GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGlnLeuGlyLysSerGluMet 100
DB 356 GAGCTCCACTGGGGCCAGAGTCCCACTCATCTGCCCCGAGCTGGGAAAGTCAGAAATG 415

QY 101 GluPheIleGluSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 416 GAGTTCATTGAAGAAGACGCCCTCGCTAGAGTGTGCTGACCCCTGCTGCGACCG 475

QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 476 TCACCCCTGCTGGCCACGGGCGCAGCTGCGGATCTGAAAGACCTCACCAAGACCGTAGC 535

QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 536 CTACAGGGAGAGCTGGAAACGGGTCTCCCCCGAGCCCCCGACATGACCTCGAGCTG 595

QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 596 GAGCTGTGCGCCACAGCGCTCTCCAAAGAGAGCTATCCAAACATGACCCCGTGCAGC 655

QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200
DB 656 CGAGAAATCCATCGTAGAGACAGAGATCTTAAGCTGAAGAAAGACGACAAAGCTG 715

QY 201 GlnGluGlnAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 716 GAGGAGAGAGCTCCAAAGCCGCGAGAGCTGAAGAGCCGTGTACCGCGCCCATCGAG 775

QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 776 TCAGACACCGGACGCGGTGAGATCATCTACGACGAAACCGGAAAGAGGTGAAGCT 835

QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 836 GCACATCGGATCTGAAAGGCTGGGGCCCGAGGTGAGCTGCCGTGTACCAACGAGCC 895

QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
DB 896 TCCGACACCCGGCAGTATCATGAGACATCAAAATAACAGCGGATCCGGAAGACCTA 955

QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
DB 956 ATCTTGACTTCAAGAGAGAAATCAACGCTCGGAAACAAATGGAGCAGAAATTCTGCAG 1015

QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysValGluGlnAlaGlnIleGlnAsnAspPro 320
DB 1016 CGCTATGACCACTCATGAGAGCCCTGGAGAAAGAGTGAAGCGCATGAGAAACACCC 1075

QY 321 ArgArgArgAlaLysGlnSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 1076 CGCGCGGGGCCCAAGAGAGACAAAGTCCGGAATCATCAAGAAAGCACTTCCCTGATC 1135

QY 341 ArgLysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySerGly 360
DB 1136 CGCAAGCAGCGCGAGCTGCAGAGCGCATGACAGAGCGGTGGGCCAGCGGGCAGTGGG 1195

QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1196 CTGTCCATGTGCGCCCGCCAGCAGCAGAGGTGTCAAGATCATCATGATGCGCTCTCA 1255

QY 381 GluGlnGluAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1256 GAGCAGAGAACTTGAGAAAGCAATGCCGCACTGCCCTGATCCGCCCATGCTTAC 1315

QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1316 GACGCTGACAGCAGCGCATCAAGTTTCATCAATGAACGGGCTTATGCGCACCCATG 1375

QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
DB 1376 AAGGTGTACAAACCGCCAGCATGATGAGTGAAGAGAGGAAAGGAACTTC 1435

QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGluLeuIleAlaSerPheLeuGluArg 460
DB 1436 CGGAGAAAGTTCTGACGATCCCAAGAACTTGGCTGATCCCATCTCTCGAGAGG 1495

QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1496 AAGACAGTGGCTGAGTGGCTCTCTATTACTGACTGACCTGAAGAAATGAAGACTATTAAG 1555

QY 481 SerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1556 AGCCTGTGAGAGGAGCTATCGGCGCCGCGCAAGACGACGACAAACAGCAGCAG 1615

QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
DB 1616 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1675

QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValGluAsn 540
DB 1676 GATGAG 1735

QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerLysGluAsnAspGlu 560
DB 1736 GACAAAGAAAGCTCTCTCAAGAGAAACAGACAGACCTCAGAGGAGGACCAAGCAG 1795

QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgGlyLysGly 580
DB 1796 AAGAGAGCTGTGCTCCAAAGCCCAAACTGCCAACAGCCAGGAGAGACCAAGAGC 1855

QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1856 CGCATCACCGCGCTCATATGAGGCAACAGAGAGAGAGGCGCATCACCCCGCAGCAG 1915

QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMet 620
DB 1916 AGCGCGAGCTGCTCCATGAGCTGAATGAGATTCTCGCTGACAGAGAAAGAAATG 1975

Qy	621	GIuThrAlaIySblySGIyLeuLeuGIuHlSGIyArgaMnTPSeRAlAlleAlaIyMet	640
Db	1976	GAACACACCACCAAGAAAGTCTCTCGGAACACGGCCGCAACTGTGTGGCCATGGCCCGGATG	2035
Qy	641	ValGIySerIySThRValSerGIyCybIySaSnPheTyPheAmTyLyLeYsArgGIu	660
Db	2036	GTGGGCTCCAGAGACTGTGTCCGACTGTGAGAACTTCTACTTCACTACAGAAAGGCAG	2095
Qy	661	AsnLeuAspGIuIleLeuGIuGlnHlSIyLeuIyMetGIuLySGIuArgAsnAlaArg	680
Db	2096	AACTCGATGAGATCTTTCAGACACACAACTGTAAGATGAGAGAGAGAAAGCAAGCGCGG	2155
Qy	681	ArgIySIySIySIySIySIyAlaProAlaAlaIaSerGIuGIuAlaAlaPheProProVal	700
Db	2156	AGGAAGAAAGAAAGAAAGCGCCGCGCGCGCCAGCAGAGGGCTGTGATTCGCCCGGTGTG	2215
Qy	701	GIuAspGIuGIuMetGIuAlaSerGIyValSerGIyAsnGIuGIuGIuMetValGIuGIu	720
Db	2216	GAGGATGAGAGATGAGAGGCTCCGGGCTGTGACGGAAATGAGAGAGAGATGGTGAAGAG	2275
Qy	721	AlaGIuAlaLeuHlSaIaSerGIyAsnGIuValProArgGIuIyCySserGIyProAla	740
Db	2276	GCTGAGGCTTACATGCTCTGTGGGAATGAGGTGCCAGAGGGGAATGACATGGCCGACGC	2335
Qy	741	ThyValAsnAsnSerSerAspThrGIuSerIleProSerProHlSThrGIuAlaAlaIyS	760
Db	2336	ACTGTCAACAAACACTCAGACACGAGAGCATCCCTCTCTCTCACTCAGAGCGCCAGAG	2395
Qy	761	AspThrGIyGlnAsnGIyProIySProProAlaThrLeuGIyAlaAspGIyProProPro	780
Db	2396	GACACAGGGCAGATGGGCCCCAGGCCCCACACCTTGGGCGCCGACGGGCAACCCCA	2455
Qy	781	GIyProProThrProPro-ArgArgThrSerArgAlaProIleGIuProThrProAlaSe	800
Db	2456	GGGCAACCCACCCACACCGAGAGACATCCCGGC-CCCACTAGGCCACCCGGGCTC	2514
Qy	800	rgIuAlaThrGIyAlaProThrProProProAlaProProSerProSerAlaProProPr	820
Db	2515	TGAAGCCACCGGAGCCCTTACGCCCCCAACGACCCCATTCGCCCTCTGTGACCTCTCC	2574
Qy	820	oValValProIySbGIuGIySbGIuGIuThRAlaAlaIaProProValGIuGIuGI	840
Db	2575	TGTGTCTCCCAAGAGAGAGAGAGAGAGAGACCGACAGACGCCGCCCAAGTGAAGAGAGG	2634
Qy	840	YGIuGIuGIuIySbProProAlaAlaGIuGIuLeuAlaValAspThrGIyLySaIaGIuGI	860
Db	2635	GGAGAGACAGAAAGCCCCCGCGCTGAGAGACTGGCACTGGACACAGAGGAAGCCGAGGA	2694
Qy	860	uProValIySbSerGIyCySThrGIuGIuAlaGIuGIuGIyProAlaIySIyLySaSPAl	880
Db	2695	GCCCGTCAAGAGCGAGTGCAGGAGAAAGCCGAGAGGGGCGCGGCAAGGCAAGAGACGC	2754
Qy	880	aGIuAlaAlaGIuAlaThRAlaGIuGIyValaIleuIySaIaGIuIySIySbGIuGIyGIySe	900
Db	2755	GGAGGCGGTGAGGSCACAGGCGCGAGGGGCGCTCAAGCACAAGAAAGAGGGCGGGAG	2814
Qy	900	rgIyArgAlaThrThRAlaIySbSerGIyAlaProGIuAspSerAspSerSerAlaTh	920
Db	2815	CGGAGAGGCCACCAAGCAGCAAGCAAGACTGGGGGCGCCCCCGAGACAGCACTCAATGCTAC	2874
Qy	920	rCySerAlaAspGIuValaAspGIuAlaGIuGIyGIyValSIySaSnArgLeuLeuSerPr	940
Db	2875	CTGCAGTGCACAGCAGGTGATGAGAGCGGAGAGCGGCGACAAGAACCGGCTGTGTCTCC	2934
Qy	940	oArgProSerLeuLeuThrProThrGIyAspProArgAlaAsnAlaSerProGIuIySPr	960
Db	2935	AAAGCCCAAGCTCTTCAACCCGACTGGGAGACCCCGGGCCAAATGCTTCAACCCAGAAAGC	2994
Qy	960	oLeuAspLeuIySGIuLeuIySGIuArgAlaAlaAlaIleProProIleGIuValThIy	980
Db	2995	ACTGGAAGCTGAAGAGACTGTAAGACAGCAGCGGCTGTCCATCCCCCCCATCCAGGTACCA	3054
Qy	980	sValHlSGIuProProArgGIuAspAlaAlaProThRySbProAlaProProAlaProPr	1000

Dd	3055	AGTCATGAGCCCCCGGGAGAGCGAGCTCCACCAAGCAGCTCCCCAGCCCAAC	3114
Qy	1000	oProProGlnaBnLeuGInProGluSeAaspAlaProGInaBProGlySerSerProAr	1020
Dd	3115	GCACCCGGAAAACCTGCAGCGCGGAGAGCGAGCCCTTCAGAGCGCTGGCAGCAGCCCG	3174
Qy	1020	gGlyLySerSerArgSerProAlaProProAlaAspLyGluLAlaPhaLAlaGluAlaG	1040
Dd	3175	GGGCAAGAGCGAGAGCCCGGAGACCCCGCGCGACAAAGAGGCTTTCAGCGAGGCCCA	3234
Qy	1040	nLyLeuBProGlyAAspProProGlySTpThSerGlyLeuBProPheProValProProAr	1060
Dd	3235	GAACTGCTTGAGGAGCCCCCTTCTGTGACTTCCGGCTGCTCTTCCCGTCCCGCCCG	3294
Qy	1060	gGlyValLLeLySaLaserProHISAlaProAspProSerAlaBaserTyraLalProPr	1080
Dd	3295	TGAGGTGATCAAGGCTCTCCCGCATGCGCGGAGCCCTCAGCTTCTCTAAGCTTCAC	3354
Qy	1080	oGlyHISProLeuBProLeuGlyLeuHISAspThLAlaGProValLeuProAlaGProPr	1100
Dd	3355	TGTCACCCCATGCGCCCTGGGCTTCATGACACGCCCGGCGCTCTGCGCGCCAC	3414
Qy	1100	oThLLeSerAnProProProLeuLISerSerAlaLyHISProSerValLeuGluAr	1120
Dd	3415	CACCATCTCCAAACCGGCTCCCTTCATCTCTCTGCCAAGCAGCCCAAGCTCTCGAGAG	3474
Qy	1120	gGlnLLeGlyAlaLISerGInGlyMetSerValGlnLeuHISValProTySerSerGluH	1140
Dd	3475	GCAATATGATGCCATCTCCCAAGGAATGTGGTCCAGCTCCAGCTCCGATCTCAGAGCA	3534
Qy	1140	sAlaLySaLProValGlyProValThrMetGlyLeuBProLeuBProMetLAspProLyGly	1160
Dd	3535	TGCCAAGGCCCCCGGTGGGCTCTGCACATGGGGCTCCCTGCCATGAGCCCAAAAA	3594
Qy	1160	sLeuAlaBProPheSerGlyValLySGInGluLeuSerProArGlyGlnAlaGlyPr	1180
Dd	3595	GCTGCACCCCTTCAGCGAGTGAAGACAGAGCAAGCTGTCCCAAGGGGCCAGGCTGGAGCC	3654
Qy	1180	oProGlySerLeuGlyValProThraLInGInGluAlaSerValLeuArGlyThraLale	1200
Dd	3655	ACCGGAGAGCTGGGGGGTGCACAGCCAGAGAGGCGTCGTGTGAGAGGAGCAAGCTCT	3714
Qy	1200	uGlySerValProGlyGlySerLLeThrLyGlyLleProSerThraLyAlProSerAs	1220
Dd	3715	GGGCTCAGATTCCGGGCGGAGCATCACCAAGCATTTCCAGCACAGGGTGGCTTCGGA	3774
Qy	1220	pSerLAlaLLeThrTyraArgGlySerLLeThrHISGlyThProLAlaBArValLeuTyLy	1240
Dd	3775	CAGGCGCATCAATACCGGGCTCATCACCAAGCAGCAGCGCAGCTAAGCTCTGTACAA	3834
Qy	1240	sGlyThrLLeThrArgLLeLISegLyGluAAspSerProSerArgLeuAspArgLyArgL	1260
Dd	3835	GGGACCATCAACAGAGTATCGGCGAGAGCAGGCCAAGTCGTGGAACCGGCGCGGGA	3894
Qy	1260	uAspSerLeuBProLySGlyHISValLLeTyGluGlyLyLySgLyHISValLeuSerTy	1280
Dd	3895	GGACAGCGCTGCCCAAGGGCCACGTCATCTCAAGAGGCAAGAGGCGACGCTGTGTCTA	3954
Qy	1280	rGluGlyGlyLysSerValThrGInCySerLySgLyuAspGlyArgSerSerSerGlyPr	1300
Dd	3955	TGAGGGTGGCATGTCTGTGACCCAGTCTCCAAAGAGAGAGGCGAAGAGAGCTAGAAC	4014
Qy	1300	oProHISGluThraLAlaLProLySArGThrTyraAspMetMetGlyLyArgValGlyAr	1320
Dd	4015	CCCCCATGAGAGCGCGCGCCCAAGCGCAGCCTTAATGACATGATGAGAGGGCGCGTGGCAG	4074
Qy	1320	gAlaLLeSerSerAlaSerLLeGlyGlyLeuMetGlyArgAlaLLeBProProGluArgH	1340
Dd	4075	AGGCATCTCTCAACCGAGCATCGAAGGTCTCATGGCGGTGCATCCCGCGAGGACACA	4134
Qy	1340	sSerProHISLLeuLySGluGlnHISValLeArgGlySerLLeThrGInGlyLlePr	1360

Db 4135 CAGCCCCACCACCTCAAGAGCAGCACCAATCCGGGGCTCATCACAGAGGATCCC 4194
Qy 1360 oATSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaIleuLeuLysPsr 1380
Db 4195 TCGGTCTTACGTGAGGACAGAGAGACTACTCTGCTCGGAGGCCAAAGCTCTTAAAGCG 4254
Qy 1380 gGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrIleThrGly 1400
Db 4255 GAGAGGACACCTCCGGCCCCACCGGCTCAAGGAGCTGACCGAGGCTTCAAGAGCGCA 4314
Qy 1400 nAlaIleuGlyProleuLysLeuLysProAlaIleGluGlyLeuValAlaThrValIlyVgl 1420
Db 4315 GGGCCCTGAGGCCCCCTGAAGCTGAAGCGCGCCCATGAGAGGCTGTGGCTCAAGTGAAGGA 4374
Qy 1420 uAlaGlyArgSerIleIleGluIleProArgGluGluIleuArgHisThrProGluLeuPsr 1440
Db 4375 GGGGGGGCGCTCATTCATGAGATCCCGCGGAGAGACTGCGGACACAGCCCCAGCTGCC 4434
Qy 1440 oLeuAlaProArgProleuLysGluGlySerIleThrGlnGlyThrProleuLysTyrAs 1460
Db 4435 CTTGGCCCCCGGGCGCTCAAGAGAGGGCTCATCACAGAGGACCCCGCTCAAGTACGA 4494
Qy 1460 pThrGlyAlaSerThrThrGlySerIlyblybHisAspValArgSerIleGlySerPsr 1480
Db 4495 CACCGGCGCTTCCACCACTGGCTTCAAAAGCAAGAGTACGCTCCCTCATCGGCAGCCCC 4554
Qy 1480 oGlyArgThrPheProProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGly 1500
Db 4555 CGGGCGGACCTTCCACCCGCTGACCCGCTGAGATGATGAGCGACCCCGGGGACTGGA 4614
Qy 1500 uArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db 4615 ACGGCTGCTGACGAGAGAGCGCTGAAGAGCGGCGCAGAGGACCGCCAGCAAGCTCGGGGG 4674
Qy 1520 ySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPsr 1540
Db 4675 CTCATTGGCGCGGCGCCCGCTCATTTGCTGAGCTGAGCTGGCGCAAGCGCGCGAGAGCCC 4734
Qy 1540 oLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVa 1560
Db 4735 CTTAACCTTATGAGGACCAAGGGGACCTTTGGCGGACCTTCCAGAGAGTTGCGCCGT 4794
Qy 1560 lThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerIlyAlaSe 1580
Db 4795 GACCAAGCGGAGGCCACCGCGCGCTGACGAGAGGAGCCTTTGCTTCCAGCAAGGATC 4854
Qy 1580 rGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrVa 1600
Db 4855 CCAAGAACCGAAAGCTGACGTGACGCTCGTGAATGCCAAAGTCCCGGACAGCACCGT 4914
Qy 1600 lProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGly 1620
Db 4915 GCGCGAGACACCAACCAACCTTCTGCTTATGACACCTGCTTGGGGCGGAGTGG 4974
Qy 1620 yValAspLeuTyrArgSerHisIleProleuAlaPheAspProThrSerIleProArgGly 1640
Db 4975 CGTGAACCTGTATGAGGACCAATCCCGCTGGCTTGAACCCACCTCCATACCCCGCGG 5034
Qy 1640 yIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAspProTh 1660
Db 5035 CATCCCTCTGAGAGCGCCGCTGCTACTACTGCCCCGACACTGGCCCCCAACCCAC 5094
Qy 1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGly 1680
Db 5095 CTACCGGACCTTATCCACCTTATCTCATTCGGGGCTACCCGAGCACGGCGGGCGCTGGA 5154
Qy 1680 uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db 5155 GAACCGGACAGCATCATCATGACTCATCATCACTGACAGATGACCAACAACGGCGC 5214
Qy 1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSerIle 1720
Db 5215 CACCGCATGGCCAGAGAGTATATGTGAGAGGGGCTCTCGCCCGCAAGTCTCGCT 5274

Qy 1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisIle 1740
Db 5275 GGCACCTAACTACGCTCGGGGTCCCGAGGACATCATGACTGTGCTCCAAAGTCCACACT 5334
Qy 1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
Db 5335 GCGTGTCTGTGCTCCCGACACAGGACCCCAAGCCGACCGCTGAGACCGCTTGGCTTA 5394
Qy 1760 rLeuProThrAlaProGlnProPheSerArgHisIleSerSerProleuSerProGly 1780
Db 5395 CTTCCACCGCGCCCTTCAAGACCTTCAAGACCCGCAAGAGAGCTCCACCTTCCAGG 5454
Qy 1780 yGlyProThrHisLeuThrLysProThrThrThrThrSerSerSerGlnArgGluArgAspAr 1800
Db 5455 AGGTCCAAACACTTGAACAACAACAACAACAGTCTCTCTCGAGGGGAGGAGAGCGG 5514
Qy 1800 gAspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThr 1820
Db 5515 GATTCGAGAGCGGAGACGGGATCGGAGCGGAAAGTCCATCTCAGTCCACACAGAC 5574
Qy 1820 rValGluHisAlaProIleThrArgProGlyThrGlnSerSerGlySerSerGlySe 1840
Db 5575 GGTGAGACAGCACCACTTGAAGACTGTGTAAGAGACAGAGAGAGCGGACAGCGGAG 5634
Qy 1840 rSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5635 CACCGCGGGGGTGGGGGACAGACAGCCGCCCCGCTCCCATCTCCATCCACACAGCA 5694
Qy 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAs 1880
Db 5695 CTCGCCATCTCCCTCGAGCCAGAGTGCCTTCCAGACAGACCCAGTGTGCTTCAAA 5754
Qy 1880 nThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuAspSe 1900
Db 5755 CACAGGATGAAGGATATCATCACCGCTGAGGCCACAGACCGCACCGTCTCGAGTGC 5814
Qy 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPsr 1920
Db 5815 CACTTCACCTTCTTACCCGCTTGGCCGCGGTGCACATTCACCTCCACCTCCACTGCCC 5874
Qy 1920 oLeuGlyGlyThrLeuAspGlyValTyrProThrLysMetGluProValLeuLeuProLy 1940
Db 5875 ACTGGGCGGACCCCTGATGGGTCTTACCTTACATGAGACCCGCTTGGCGCCAA 5934
Qy 1940 gGluAlaProArgValAlaArgProGluArgProAlaGlyAlaAspThrGlyHisAlaPheLe 1960
Db 5935 GAGAGCCCCCGGGTGGCGGACAGAGCGGCCGAGAGACCGGCATGCTTCCCT 5994
Qy 1960 uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGly 1980
Db 5995 CGCCAAAGCCCCAGCCGCTCGGGGCTGAGAGCCGCTCTTCCCGACAGAGGGCTCGGA 6054
Qy 1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db 6055 GCGCGGCGCTTGTGCTGCTCTGTCTGCGCAGCGCACCATGCGCGCACCCCTGGGA 6114
Qy 2000 sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerArg 2020
Db 6115 GAACCTGCACTTCAACAGCGAGCGCGGAGCGCGGCGGCGCACTGCTCGGCTCGGA 6174
Qy 2020 pProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLysSe 2040
Db 6175 CCGGACCGGAAAGAACTCAAGTAAACCTTTTTCATTCAGAGAACTGAACTCCGTTTC 6234
Qy 2040 rLeuGlyTyrHisGlySerSerTyrSerProGlnGlyValGluProValSerProValSe 2060
Db 6235 TCTGGGTTTACAGCGAGAGCTTACAGCCCGGAAGGGGTGAGACCCGTGAGCCTGTGAG 6294
Qy 2060 rSerProSerLeuThrHisAspLysGlyLysProLysblybHisLeuGlnGluLeuAspLysSe 2080
Db 6295 CTCACCACTGTACCAAGACAGCAAGGGGCTCCCAAGACCTGGAAGAGCTGACAAAGAG 6354

QY	2080	HHIALENGUGLYGLIUEUAAGPProLYSGINProGLIProVALyLyeNGUYGLYGLIUA1	2100
Db	6355	CCACCTGGAGGGGGAGCTGGGGCCCAAGCACCGACGCCCTGTGAAGCTTGCGGGGAGGC	6414
QY	2100	AAIAHISLEUPROHISLEUAAGPProLEUPROGLISERGInPROSeSERSeSERProLEULe	2120
Db	6415	CGCCCACTCCCAACACCGGGGGCGGTGGCTGAGAGCGCAGCTGTCTACACCGCGTGTCT	6474
QY	2120	UGINThRALAPROGLYVALYLSGLYHIEGLINARGVALYALThIRLEUAlAGINHISLIESe	2140
Db	6475	CCAAAGCCGCCCAAGGGGTCAAAAGGTACACACGGGTGTGTACCTGTGGCCCGACATCAG	6534
QY	2140	RGULVALILethRGInAPRYThRTARGHISIHISProGLInLIneUSERAlAPROLeUPR	2160
Db	6535	TGAGGTCACTCACAGAGACTACCCGGCGACCCACCAAGAGCTCAGCGCACCCCTGTCC	6594
QY	2160	OALAPROLeUPRYSeRPhePROGLYALASERCySProVALIeUAAPLeUAAGPProPR	2180
Db	6595	CGCCCCCTCTACTCTCTTCCCTGGGGCCAGCTGGCCGTCTTGAGACTCTCGCGCCACCC	6654
QY	2180	OSERASPLEUPYrLEUPROProPROASPHISGLYALAPROAlAARGLISERPROHISSE	2200
Db	6655	CAGTGACCTTACCTCCGGCCCCCGGACCATGTGTGCCCGGCGCGTGGCTCCCCCACAG	6714
QY	2200	RGULGLYGLYLYSARGSERProGLINProANLyeThSERValIeNGUYGLYGLIUASe	2220
Db	6715	CGAAGGGGGCAAGAGCTCTCCAGAGCCAAACAGCGTGGTCTTGGTGTGTGAGGA	6774
QY	2220	PGLYILEGLUPROVALSERProPROGLUGLYMETHRGInUPROGLYHISERASGERAl	2240
Db	6775	CGGATTTGAACCTGTGTCCCAACCGAGGGCATGACGAGCGCAGGCACTCCGGAGTGC	6834
QY	2240	AVALTYrPROLeUeUENTYrARGASPGLYGLUGINTHRGInUPROSeRARGMETGLYSerLY	2260
Db	6835	TGTGTACCCGCTGCTGTACCGGGATGGGGAAACAAGCGGACCCAGACAGATGGGCTCCA	6894
QY	2260	BSERPROGLYANThRSERGInProPROAlAPhePheSERLYSLEUTHRGInUSERANSe	2280
Db	6895	GTCTTCACGGCAACACAGCCAGCGCCGCTTCTTCGCAAGCTGACCGAGAACATCTC	6954
QY	2280	RAIMETVALLYSSErLYSLYSGINGLIILASNLySLYSEUANTHRHISANARGAS	2300
Db	6955	CGCCATGGTCAAGTCCAAAGAACAGAAATCAACAAGAGCTGAACACCCACACCGGAA	7014
QY	2300	NGInUPROGLYrASNLIESERGInPROGLYThRGInULIePheANMeProAlALETH	2320
Db	7015	TGAGCTGTAATACATAATACACCGACCTGGGACGGAGATCTTCAATATAGCCCGCATAC	7074
QY	2320	RGLYThRGlyLEUeMETHrTYrARGSERGLINALAYAlnGLUHISAlASERThRASMe	2340
Db	7075	CGGAACAGGCGCTTATGACTTATAGAACCCAGCGGTGCAGGAACATCGCACACCAACAT	7134
QY	2340	TGLYLeNGInLAlIeLLeARGLYSAlALEUeMGLYrSYrYrASPGINTRPGLUGInUSE	2360
Db	7135	GGGGCTGAGGCCAATATTAAGAAAGCACTATGGTAAATATGACCGTGGAGAGATC	7194
QY	2360	RPROProLeUSERAlASenAlAPheASNPProLeUeSNAlASERAlASERLeUPROAlAAl	2380
Db	7195	CCCCCGCTCAGCGGCATGCTTTTAACCTCTGAATCCAGTCCAGCTGCCCGCTGC	7254
QY	2380	AMEePROLIEThRALAAlASPGLYARGSERASPHISThIRLeUTHRSERProGLYGLYGLI	2400
Db	7255	TATGGCCCTTAACCGCTGTCTGACGGACGGAGTGCACACACTCACCTCGCCAGGTGGCG	7314
QY	2400	YGLYLYSAlALySVALSERGLYARGProSERSeRARGLYSAlALySBRPROAlAPROGL	2420
Db	7315	CGGGAAGGCCAAGGCTCTGTGCAGCCAGAGCGGAAAGCCAAAGTCCCGGCGCCGG	7374
QY	2420	YLEUAlASERGlyASPARGPProPROSeCYALISERSERVALHISSErGLUGLYASPCYAS	2440
Db	7375	CCTGGCATCTGGGAGCCGGCAACCTCTGTCTCTCTCACTGACACTCGAGGGAGACTGCA	7434
QY	2440	NAIRGARGThRPROLeUTHRASnARGVALITRPGInASPARGPROSeSERAlAGLYSeTh	2460

Db	7435	CCGCGGAGACCGCGCTACCAACCGCGTGTGGAGAGACAGGCGCCCTCGTCCGAGGTTCCAC	7494
QY	2460	RPRPheProTyrAsnProleuileMetArgLeuGlnAlaGlyValMetLAsEProPr	2480
Db	7495	GCAATTCCTCCATCAACCCCTGATCATAGCGGTGACAGCGGGGTGCATGGCTTCCACC	7554
QY	2480	oProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaTrpAspG1	2500
Db	7555	CCCAACGGGGCTCCCGCGGGACAGGGGGCCCTCGCTGGCCCCACCAACCGCTGGAGCA	7614
QY	2500	uGlnProLysProLeuLeuLeuGlySerGlnTyrGlnThrLeuSerAspSerGlu	2517
Db	7615	GGAGCCCAAGCACTGCTGTCTGCAGTACGAACACTCTCCGACAGCCGAG	7666

```

RESULT 5
US-09-819-104A-3
; Sequence 3, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: IMG-030
; CURRENT APPLICATION NUMBER: US/09/819,104A
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7521
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(7521)
US-09-819-104A-3

Alignment Scores:
Pred. No.: 0 Length: 7521
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 28
Query Match: 98.21% Indels: 2
DB: 10 Gaps: 3

US-09-522-753-5 (1-2517) x US-09-819-104A-3 (1-7521)
QY 1 MetSerGlySerThrGlnLeuValAlAGlnThrTPyrGalaThrGluProArgTyrPro 20
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 ATGTGGGGCTCCACACACCCTGTGGCACACAGCGTGAGGGGCATGAGCCCCGTACCCG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisTrpAspValGlyLeuLeu 40
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 CCCACAGACCTTTTCCTACCCGATGCCAGATGCCCGGACGACACAGCATGTCGGCTCCTG 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 41 GluIlyrGlnHisHisSerArgArgAspTyrAlaSerHisHisLeuSerProGlySerTlleIleGln 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 GAGTACCAAGCACCACTCCCGGACTATAGCCTCCACCTGTGGCCCGGCTCATATCCAG 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 ProGlnArgAlaArgPProSerLeuLeuSerGluPhcGlnProGlyAsnGluArgSerGln 80
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 CCCAGAGGGGGGAGAGCCCTCCCTGCTGTGAATTCCAGCCCGGAAATGAACGGTCCACG 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 81 GluLeuHISLeuArgPProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 GAGCTCCACCTCGCGCCAGAACTCCCACTCAATACCTGCCGAGCTGGGGAATCAAGAATG 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuAlaArgPro 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 GAGTTTCATTAAGAACAGACGCCCTTCGGCTAGAGCTGCTGCTGTAGACCCCTGTGTCGACCG 360

```

QY 121 SerProLeuLeuValaThrglyGlnProAlaGlySerGluAspLeuThrTlyAspArgSer 140
Db 361 TCACCCCTGTGGCCACGGGCCAGCTGGGGATCTGAAGACCTCACCAAGAACGTAAGC 420
QY 141 LeuThrGlyTlybLeuGluProValSerProPProSerProProhiIstThrAppProGluLeu 160
Db 421 CTGACGGGCAAGCTGGAAACGGGTGTCTCCCGCACGCCCGGCACACTGACCTCGAACCTG 480
QY 161 GluLeuValProProArgLeuSerTlybGluLeuLeuLeuAsnMetAspArgValAsp 180
Db 481 GAGCTGGTGGCCGACCGCTCTCCAAAGAGAGCTGATCCGAACATGACCGCGCTGAGC 540
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerTlybLeuTlybGluGlnGlnLeu 200
Db 541 CGAGAGATCAACATGTAGAGACAGATCTTAAGCTGAAGAAAGAACAGCAACAGCTG 600
QY 201 GluGluGluValAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
Db 601 GAGAGAGAGCTGGCAAGCGCGCCGAGCTGAGAAAGCCGTGTACCGCGCCCATCGAG 660
QY 221 SerTlybIleArgSerLeuValGlnIleIleTyrArgGluAsnArgLysbValaGluVala 240
Db 661 TCGAAGACCGGACCGCTGTGACAGATCTACGACGAAACCGGAAGAAAGCTGAAGCT 720
QY 241 AlaIleArgIleLeuGluGluTlyLeuGlyProGlnValaGluLeuProLeuTyrAsnGlnPro 260
Db 721 GCACATCGATTTCTGAAAGGCTGGGGCCCGAGGTGAGGTGCGGCTGTCAACCAACCCC 780
QY 261 SerAspThrArgGlnTyrIleGluAsnIleLysIleAsnGlnAlaMetArgLysbLeu 280
Db 781 TCCGACACCGGCAAGTATCATGAAACATCAAAATAAACAGCGGATCGGAGAAAGACTA 840
QY 281 IleLeuTyrPheLysArgArgAsnIleAlaArgLysGlnTyrbLysGlnLysPheCysGln 300
Db 841 ATCTTGACTTCAAGAGAGAGATCACGCTGGAAACATAGGAGCAAGAGTTCTGCGAG 900
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysbLysValaGluArgIleGluAsnAsnPro 320
Db 901 CGCTATGACCAAGCTCATGAGGCTGGAGAAAGAGTGGAGGCATCGAAGAACACCC 960
QY 321 ArgArgArgAlaLysGluSerLysbValaArgGluTyrTyrGluLysGlnPheProGlnIle 340
Db 961 CGCGCGGGGCCAAGAGAGCAAGGTTGCGAGTACTACAGAAAGCGATGTTCCCTGAGATC 1020
QY 341 ArgLysGlnArgGluLeuGlnGlnIleArgMetGlnSerArgValaGlyGlnArgGlySerGly 360
Db 1021 CGCAAGACCGCCAGGCTGCGAGAGCGCATGCAAG--AGGTTGGGCGCGCGGAGTGGG 1077
QY 361 LeuSerMetSerAlaAlaArgSerGluIleGluValSerGluIleIleAspGlyLeuSer 380
Db 1078 CTGTCCATGTGCGCCGCGCGCGAGCAAGAGTGTCAAGATCANTGATGCGCTCTCA 1137
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 1138 GAGCAGAGAACTCGGAGAGAGATGCGCACCTGGCGCTGATCCCGCCCATGCTGTAC 1197
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db 1198 GAGCGTGAACAGAGCGCATCAAGTTCTATCAACATGAACGGGCTTATAGCGGACCCCATG 1257
QY 421 LysValTyrLysAspArgGlnValaMetAsnMetTyrSerGluGlnGluLysGluThrPhe 440
Db 1258 AAGGTGTACAAACACCGCCAGTCAATGAATGTGAGTGAAGAGAGAAAGAGACCTTC 1317
QY 441 ArgGluLysPheMetGlnIleProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 1318 CGGGAAGAGTTCAAGACATCCCAAGACTTGGCTGTATCCCATCATCTCGAGAGG 1377
QY 461 LysThrValaIaGluCysValaLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db 1378 AAGACAGTGGCTAGTGGCTCTCTATTACTTACTGACTAAGAAAGAAATGAATCTTAAG 1437
QY 481 SerLeuValaArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500

Db 1438 AGCTGTGAGAGCGAGACTATCGCGCGCGCAAGGCCAGCAACAAACAGACAG 1497
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
Db 1498 CAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
Db 1558 GATGAG 1617
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyIaAspAsnAspGlu 560
Db 1618 GACAAAGAAACCTCTCAAG 1677
QY 561 LysGluAlaValaIleSerLysGlyArgLysThrAlaAsnSerGlnIleArgLysGly 580
Db 1678 AAGAGGCTGTGCTTCCAAAGGCGCCAAATCTGCCAAACAGCGGAGAAACCGCAAGGC 1737
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1738 GCGATCACCGCTCATGCTTAATGAGCCAAACAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
Db 1798 AGCGCGAGCTGGCTTCATGAGAGTGAATGAGAGTCTCGCTGACAGAAAGAAATG 1857
QY 621 GluThrAlaLysLysGlyLeuLeuGluLysGlyArgAsnTyrSerAlaIleAlaArgMet 640
Db 1858 GAAACAGCCCAAGAAAGTCTCTCGGAACACGCGCGCAACTGTGCGCATCGCCGATG 1917
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheThrPheAsnTyrLysbValaArgGln 660
Db 1918 GTGGGCTCCAGACTGTGTGCGAGTGAACCTTCACTTCAACTCAAGAAAGAGCAG 1977
QY 661 AsnLeuAspGluIleLeuGlnGlnIleLysbLeuLysMetGluLysArgAsnAlaArg 680
Db 1978 AACCTCATGATGATCTTTCAGAGCAGCAACAGCTGAATGAGAGAGAGAGAGAGAGAG 2037
QY 681 ArgLysLysbLysbAlaProAlaAlaIleSerGluGluAlaIlePheProProVala 700
Db 2038 AGGAAAGAAAGAAACCGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2097
QY 701 GluAspGluGluMetGluLysSerGlyValaSerGlyAsnGluGluMetValaGluGlu 720
Db 2098 GAGATGAGAGATGAGAGGCTGCGCGTGAACGGAATGAGAGAGATGTGAGAGAG 2157
QY 721 AlaGluAlaLeuIleAlaIleSerGlyAsnGluValaProArgGlyGluCysSerGlyProAla 740
Db 2158 GCTGAA-----GCC 2166
QY 741 ThrValaAsnAsnSerSerAspThrGluSerIleProSerProhiIstThrgluAlaAlaLys 760
Db 2167 ACTGTCAACAAACGCTTCAGACACCGAGAGATCTCTCTTCACTGAGGCGCGCAAG 2226
QY 761 AspThrGlyGluAsnGlyProLysProProAlaIleThrGluValaIleAspGlyProProPro 780
Db 2227 GACACAGGAGAGATGAGCGCCCAAGCGCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2286
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaIle 800
Db 2287 GGGCCACCCACCCACACCGAGAGATCCCGGC-CCACAGAGATCCACCCCGGCTC 2345
QY 800 TgluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2346 TGAAGCACCTTAAGCCCTTAAGCGCCCAACAGACACCCCATTTCCCTTCACTCTCC 2405
QY 820 oValaIleProLysGluGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 840
Db 2406 TGTGGTCCCAAG 2465
QY 840 YGluGluGluLysProProAlaAlaGluGluLeuValaIleAspThrGlyLysbAlaGluGlu 860

Dh 2466 GGAGAGCAGAAAGCCCCCGGCTGAGGAGCTGCAGTGTGACACAGAGGAGCCGAGAA 2525
Qy 860 uProVallylSerGluCysThrGluGluValIleuGluGluYProAlaIleuGlyLysAspAl 880
Dh 2526 GCCCGTCAAGAGGAGGATGCGAGGAGAGCCGAGAGGGCCGAGCAAGGAGCAAGACCG 2585
Qy 880 aGluValaAlaGluValaThrAlaGluValaIleuLysValaGluLysLysGluGluYglYse 900
Dh 2586 GGAGGGCCGCTGAGAGCCAGCGCCGAGAGGGCGCTCAAGCCAGAGAGAAAGAGAGGGCGGAG 2645
Qy 900 rGlyValaGluAlaThrThralaIleuSerSerGlyValaProGluAspSerAspSerSerAlaTh 920
Dh 2646 CGGACAGGGCCACCAAGCCAGCAAGAGCTCGGGCGCCCGCCAGAGACGAGCTCCAGTCCAC 2705
Qy 920 rCysSerAlaAspGluValaAspGluValaGluGluYglYAspLysAspArgLysLeuSerPr 940
Dh 2706 CTGCACTGCAGACAGAGTGGATGAGGCTCGAGGGCGGCGACAGAACCGGGCTGTGTCTCCC 2765
Qy 940 oArgProSerLeuLeuThrProThrGlyYAspProArgAlaAspAlaSerProGluLysPr 960
Dh 2766 AAGGCCACAGCTCTCCACCCCGACTGGCGACCCCGGGCCAAATGCTCAACCCAGAAAGCC 2825
Qy 960 oLeuAspLeuLysGluLeuLysGluLysArgAlaAlaAlaIleProProGluIleGluValaThrLy 980
Dh 2826 ACTGGACCTGAAGCAGACTGAAGCAGAGCGAGCGCTGCATCCCGCCATCCAGGTCAACCA 2885
Qy 980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Dh 2886 AGTCCATGAGCCCCCGGGAGAGACGAGCTCCACCAAGCCAGCTCCCGAGCCCCCACC 2945
Qy 1000 oProProGluAsnLeuGluProGluSerAspAlaProGluInProGlySerSerProAr 1020
Dh 2946 GCCACCGCAAAACCTGACAGCCGAGAGCGACGCGCCCTCAGACAGCTGAGCAGCAGCCCCG 3005
Qy 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Dh 3006 GGGCAAGACAGAGAGCCGAGACCCCGCCGCAAGAGAGCAGAGAGCCGTGTCTT 3065
Qy 1034 ----AlpheaAlaGluValaGluLysLeuProGlyAspProProCysThrThrSerG 1052
Dh 3066 CCCAGCCTTCGAGCCGAGAGCCCAAGAACTGCTGGGGAGCCCCCTTGCTGAGCTTCCGG 3125
Qy 1052 yLeuProPheProValProProArgGluValaIleuLysAlaSerProHisAlaProAspPr 1072
Dh 3126 CCGGCCCTTCCCGTGCCTCCCGGTGAGGTGATCAAGGCTCCCGCATGCTCCCGAGACC 3185
Qy 1072 oSerAlaIlePheSerTyAlaProProGluLysIleProLeuProLeuGlyLeuHisAspThrAl 1092
Dh 3186 CTCAGGCTTCTCTTACGCTCACCCTGTGACCCACCTGCGCTGGGCTTCATACACTGC 3245
Qy 1092 aArgProValLeuProArgProProThrLysSerAspProProProLeuLysSerSerAl 1112
Dh 3246 CCGGCCGCTCTCCCGGCCACCCACATCTCCAAACCGGCTCCCTCATCTCTCTGCG 3305
Qy 1112 alybHisProSerValLeuGluValaGluValaIleGlyAlaIleSerGluYmSerValG 1132
Dh 3306 CAAGCACCCAGAGCTCTCGAGAGGCAATAATGATGTCATCTCCCAAGAAATGTCGGTCCA 3365
Qy 1132 nLeuHisValProTySerGluHisAlaLysValaProValGlyProValaThrMetGlyLe 1152
Dh 3366 GCTCCACGTCCTCTACTCAGAGCATGCCAAGGCCCGGTGGGCTGTGTCACTGAGGGCT 3425
Qy 1152 uProLeuProMetAspProLysLysLeuValaProPheSerGlyValaLysGluGlu 1172
Dh 3426 GCCCTGCGCATGAAACCCCAAAAGCTGAGCACTTTCAGCGAGTGAAGAGAGAGAGCT 3485
Qy 1172 uSerProArgGlyGluAlaGlyProProGluSerLeuGlyValaProThrAlaGluGlu 1192
Dh 3486 GTCGCCACGGGGCAGAGCTGAGCCACCGAGAGAGCTGGGGGTGCCCAAGCCAGAGAGGC 3545
Qy 1192 aSerValLeuArgGlyThralaLeuGlySerValProGlyGlySerLethrLysGlyL 1212
Dh 3546 GTCCGTGTGAGAGGAGACAGCTCTGGGCTCAGTTCCGGGCGAAGCATCCAAAGGCAT 3605

Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerLethrHisG 1232
Dh 3606 TCCAGACACAGGGGTGCCCTCGAGACGGCCATCAATACCCGGCTCATCACCCACGG 3665
Qy 1232 yThrProAlaAspValLeuTyLysGlyThrLethrAlaGluIleGlyGluAspSerPr 1252
Dh 3666 CAGCCAGCTGACGTCCTGTACAGAGGACCATTCACAGAAATCATCGGGAGAGACGCC 3725
Qy 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValaIleTyGluG 1272
Dh 3726 GAGTGCCTTGAGCCGGCGCGGAGAGACGCTGCCCAAGGGCCAGTCACTACGAAG 3785
Qy 1272 yLysLysGlyHisValaLeuSerTyGluGlyGlyMetSerValaThrGluCysSerLysG 1292
Dh 3786 CAAGAAAGGCCACAGTCTTCTCTATAGAGGTGAGCATGTGTGTGACCCAGTGTCTCAAG 3845
Qy 1292 uAspGlyAspSerSerSerGlyProProHisGluThralaAlaProLysArgThrTyTrs 1312
Dh 3846 GAGCGCAGAAACAGCTCAGAGACCCCGCATGAGAGAGCGCGCCCAAGCCGACCTATGA 3905
Qy 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerLleGluGlyLeuMetG 1332
Dh 3906 CATGATGAGGGCCGCTGGGAGAGCCATCTCTCAGCCAGCATGAGAGTCTCATGG 3965
Qy 1332 yArgAlaIleProProGluLysArgHisSerProHisHisLeuLysGluGluHisIleAr 1352
Dh 3966 CCGTGCATCCCGCGGAGCAGACAGCCCGCCACCACTCAAAAGACAGACACACATCCG 4025
Qy 1352 gGlySerLethrGluGlyLleProArgSerTyValaGluValaGluValaAspTyLeuAr 1372
Dh 4026 CGGGTCCATCAACAAGAGGATCCCTGATCTGACGTGAGGACACAGAGAGACTCACTGG 4085
Qy 1372 gArgGluAlaLysLeuLysArgGluGlyThrProProProProProProSerArgAs 1392
Dh 4086 TCGGAGAGCCAAAGCTCTTAAAGCGGAGGCAAGCTCCGCCCCACCGCCCTCAGGG 4145
Qy 1392 pLeuThrGluValaTyLysThrGluAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Dh 4146 CCGAACCGAGGCTTACAAAGACAGAGCCCTGGGCGCCCTGAAGCTGAAGCCGCCATGA 4205
Qy 1412 uGlyLeuValaAlaThrValaLysGluValaGlyArgSerLethrGluIleProArgGluG 1432
Dh 4206 GGGCTGTGTGGCACGCTGAAGAGAGGGCGGCTCATCACTGAATCCCGGCAAGGA 4265
Qy 1432 uLeuArgHisThrProGluLeuProLeuValaProArgProLeuLysGluGlySerLeth 1452
Dh 4266 GCTGGGCAACAGCCGAGACTGCCCTTGSCCGCGCGCTCAAGAGAGGCTCCATCAC 4325
Qy 1452 rGluGlyThrProLeuLysTyArgThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
Dh 4326 GCAAGGACCCCGCTCAAGTACAGACCGGCGGCTGCACCACTGCTCCAAAAAGCACGA 4385
Qy 1472 pValaArgSerLeuLleGlySerProGlyValaArgThrPheProProValaHisProLeuAspVa 1492
Dh 4386 CGTACGCTCTCCATCTGAGAGCCCGCGGAGAGGTTCACCCGAGCACCCGCTGATGT 4445
Qy 1492 lMetAlaAspAlaArgAlaLeuGluValaGlyLysTyArgGluLysSerLeuLysArgPr 1512
Dh 4446 GATGGCCGAGCGCCGGGCACTGGAAAGTGCCTGTCTACAGAGAGAGCTGAAGAGCGGGC 4505
Qy 1512 oGlyThralaSerSerSerGlyGlySerLleAlaArgGlyAlaProValaIleValaProG 1532
Dh 4506 AGGAGACCGCAGAGCTCGGGGGGCTCCATTGCGCGCGCGCCCGGTATGTGCTGA 4565
Qy 1532 uLeuGlyLysProArgGluLysSerProLeuThrTyArgLysAspHisGlyAlaProPheAlaG 1552
Dh 4566 GCTGGGTAAAGCCGGCAAGCCCGCTGTGACATATGAGGACACAGGGGACCCCTTGGCGG 4625
Qy 1552 yHisLeuProArgGlySerProValaThrMetArgGluProThrProArgLeuGluGlu 1572
Dh 4626 CCACTTCCACAGAGTTCGCCGTGTGACATGGCGGAGCCACCGCCGCTGAGAGAGGG 4685

QY 1572 ySerLeuSerSerSerlybAlaSerGlnAspArglyLeuTherTherProArgGlu1 1592
Db 4686 CAGCCTTCCTCCAGCAAGGATCCAGAGCCGAAAGTGAAGTGAAGCTGAGAT 4745
QY 1592 eAlaIbysSerProHiaserTherValProGluHiasIbProHiaserProTyrg1 1612
Db 4746 CGCGAACTCCCGAGAGCAAGCGGCGGAGACACCCCACTCCGCTTGA 4805
QY 1612 uHiasLeuArgGlyValSerGlyValAspLeuTyArgSerHiasIleProLeuAlaPh 1632
Db 4806 GCACCTGCTCGGGGCTGAGTGGCTGAGACTGTATCGAGCCCATTCCTCGGCTT 4865
QY 1632 eaSPProThSerIleProArglyIleProLeuAspAlaIaIaIaTyTyLeuPr 1652
Db 4866 GCAGCCCACTCCATACCCCGGCGCATCCCTGGAAGCGAGCGCTGCTTACCTGCC 4925
QY 1652 oAArgHiasLeuAlaProAsnProThTyTyProHiasLeuTyTyProTyTyLeuIleatrg1 1672
Db 4926 CCAGACACTGGCCCCCAACCCCACTACCCGACCTTACCCCACTTACCTCCGCG 4985
QY 1672 TyTyTyProAspThAlaIaIaLeuGluAsnArgGlnThIleIleIaAsnAspTyTyIleThrse 1692
Db 4986 CTACCCCGAGACAGGCGGCTGAGAAACCGGAGACCATCATCATATACATACCTC 5045
QY 1692 rGlnGlnMetHiasIbAsnThralaThralaMetAlaGlnArgAlaAspMetLeuArg1 1712
Db 5046 GCAGAGATGCACCAACAACAGGCGACCGGCATGGCCAGCGAGCTGATATGCTGAGGG 5105
QY 1712 yLeuSerProArgGlyLeuSerSerLeuAlaLeuAsnTyralaIaGlyProArglyIle11 1732
Db 5106 CCTCTGCCCCCGGAGTCTCTGCTGAGCTCAACTAGCTGCGGCTCCCGAGGACATCAT 5165
QY 1732 eaAspLeuSerGlnValProHiasLeuProValLeuValProProThProGlyThrProal 1752
Db 5166 CGACCTGTCCCAAGTGCACACCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5225
QY 1752 aThralaMetAspArgLeuAlaTyLeuProThralaProGlnProHiaserSerArgHias 1772
Db 5226 CACCGCATGGAGCGCTTGCCTTACCTCCCAACCGCCGCCCTTACAGACCGGCCA 5285
QY 1772 sSerSerSerProLeuSerProGlyTyProThHiasLeuThrybProThThrThrse 1792
Db 5286 CAGCAGCTCCCACTCTCCCAAGAGTCCACAACACTTGACAAACCAACCAACGCTC 5345
QY 1792 rSerSerGluArgGluArgbArgbArgbArgGluArgbArgbArgGluArgGluArgGluArg 1812
Db 5346 CTGCTCGAGAGGAG 5405
QY 1812 sSerIleLeuThSerThrThrThrValGluHiasIaProIleTyArgProGlyThrg1 1832
Db 5406 GTCCATCTCTACCTCCACCAAGCGGTGAGAGACGACCATCTGAGAGACTGTGACGA 5465
QY 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProal 1852
Db 5466 GCAGAGACAGGAG 5525
QY 1852 aSerHiaserHiasIaHiasGlnHiaserProIleSerProArgThGlnAspAlaLeuG1 1872
Db 5526 CTCCCACTCCCAAGCCCAACGACCTGCGCATCTCCCTGGAAGCCAGAGATCCCTCCA 5585
QY 1872 nglnaArgProSerValLeuHiasbAnThrgIyMetLyGlyIleIleThralaValGluPr 1892
Db 5586 GCAGAGACCCAGTGTCTTCAACAACAGGCAATGAAGGATATCATACCGCTGTGAGCC 5645
QY 1892 oSerTybProThValLeuArgSerTherTherSerSerProValArgProAlaIaIaTh 1912
Db 5646 CAGCAAGCCCAAGCTCTGAGTCACTCCACCTCTCAACCGCTTCCGACCTGAC 5705
QY 1912 rPheProProAlaThHiasCybProLeuGlyGlyThrLeuAspGlyValTybProThrie 1932
Db 5706 ATTCCCACTGCAACCACTGCACTGGGAGGAGACCTGATGGGGTCTTACCTTACCT 5765
QY 1932 uMetGluProValLeuLeuProLyGluAlaProArgValAlaArgProGluArgProar 1952

Db 5766 CATGAGAGCCCGTCTTGCTGCCAAGAGAGCCCGCGGCTGCGCGGCAAGCGGCCCG 5825
QY 1952 gAlaAspThGlyHiasIa 1972
Db 5826 AGAGAGACAGGAGCATATGCTTCTTCCGCAAGCCCGCGCTCCGAGTGGAGCCGC 5885
QY 1972 aSerSerProSerTybGlySerGluProArgProLeuValProProValSerGlyHiasI 1992
Db 5886 CTCTCTCCCAAGCAAGGAGCTCGAGAGCCCGCGCTTATGCTCTCTCTCTGCGCACGC 5945
QY 1992 aThralaIaArgThProAlaIa 2012
Db 5946 CACCATGCGCGGACCCCTGCAAGAACTCGACCTCACAGCGGAGCCGAGCCGCC 6005
QY 2012 oAlaProProAlaSerAlaSerAsnProHiasArgGlyTybThGlnSerTybProHise 2032
Db 6006 GGGCGCACTGCTCGGCTCGGAGCCGAGCCGAGAAAGACTCAAGTAAACCTTTTC 6065
QY 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTybHiasGlySerSerTybSerProGluG1 2052
Db 6066 CATCCAGAACTGAAGTCCCTTCTTGAGTTACAGGAGCACTACAGCCCGCAAG 6125
QY 2052 yValGluProValSerProValSerSerProSerLeuThHiasPlysglyLeuProLy 2072
Db 6126 GTGAGAGCCCGTCAAGCCCTGTGAGCTCACCACTGAGCCACAGAGAGGCTCCCA 6185
QY 2072 bHiasLeuGluGluLeuAspLySerHiasLeuGlyGlyLeuLeuArgProLybGlnProG1 2092
Db 6186 GCACCTGGAAGACTGCACAAAGAGCACTGGAGGGGAGCTCCGCGCCCAAGAGCCAGG 6245
QY 2092 yProValLybLeuGlyGlyGluAlaIaHiasLeuProHiasLeuArgProLeuProLyse 2112
Db 6246 CCGCTGAGACTTGGGGAGAGCCCGCACTTCCACACTTCCGCGCTGCTGAGAG 6305
QY 2112 rGlnProSerSerSerProLeuLeuGlnThralaProGlyValLyGlyHiasGlnArgVa 2132
Db 6306 CAGCCCTGCTCAGCCCGCTGTCTCCAGACCGGCCAGGGGTCAAGTCAACAGCGGCT 6365
QY 2132 yValThrybLeuAlaGlnHiasIleSerGlyValIleThrybAspTyTyThrybHiasIbP 2152
Db 6366 GGTCACTCGGCGCCACACATCATGAGGTATCAACAGGACTTACACCGGAGCCACCC 6425
QY 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyTySerPheProGlyValaSerCybP 2172
Db 6426 ACAGCACTCAGCGCACCTCTGCGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6485
QY 2172 oValLeuAspLeuArgArgProProSerAspLeuTyTyLeuProProAspHiasGlyAla 2192
Db 6486 CGTCTGAGACTTCCGCGCGCCACCAAGTACCTTACCTCCGCGCCCGGAGCATGTGC 6545
QY 2192 aProAlaArgGlySerProHiaserGluGlyTybTybArgSerProGluProAsnLybTh 2212
Db 6546 CCGGCGCGGTGCTCCCGCCACAGCAAGGGGCAAGAGTCTCCAGAGCAAAACAAGAC 6605
QY 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
Db 6606 GTGCTGTGGGTGGTGGTGGAGAGCGTATTGAACCTGTGTCCCAACCGAGGAGATTGAC 6665
QY 2232 rGluProGlyHiaserArgSerAlaValTybProLeuLeuTyArgAspGlyGlyGlnTh 2252
Db 6666 GAGGAGAGGCACTCCCGAGAGTGTGTATCCGCTGTATCCGAGAGAGAGAGAGAGAG 6725
QY 2252 rGluProSerArgMetGlySerTybSerProGlyValanThSerGlnProProAlaPhePh 2272
Db 6726 GAGGCCACAGATGGCTTCAAGTCTCAGGCAACACAGGCAACCGCCACCTTCTT 6785
QY 2272 eSerTybLeuThrgIySerAsnSerAlaMetValLybSerTybLybGlyGlnIleAsnLy 2292
Db 6786 CAGCAAGCTGACCGAGAGCACTCCGCTATGTCAAGTCCAGAGCAAGAGATCAACA 6845
QY 2292 sTybLeuAsnThHiasbAnArgbAsnGluProGluTybAsnIleSerGlnProGlyThrg1 2312

Db 6846 GAAGCTGAACACCCACACCGGAATGAGCCCTGAATACAAATATCAGCCAGCTGGAGCGA 6905
Qy 2312 uilePheanMeProAlaileThrglyThrglyLeuMeThrTyArgSerGlnAlaVa 2332
Db 6906 GATCTTCAATATGCGCCGATCAACCGGAACAGGCTTATGACCTATAGAAAGCAAGCGGCT 6965
Qy 2332 lGnGluHlSaIaSerThraSmMeGlyLeuGluAlailelLeArgLysAlaLeuMeG 2352
Db 6966 GCAGAAACATGCGACGACCAACATGCGGCTGGAGGCAATATTAAGAAAGCACTATGCG 7025
Qy 2352 ylyeTyraSPGIINTPGluGluSerProProleuSerAlaAaAlaPhaAspProleu 2372
Db 7026 TAAATATGACCAATGGGAAGAGTCCCGCGCTCAGCGCCAAATGCTTTAACTCTGAA 7085
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMeProIleThraAlaAlaSPGIYArgSerAspH 2392
Db 7086 TCCCACTGCGAGCTCCCGCTGCTATGCCCAACCGCTGCTGAGAGAGAGTACCA 7145
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyYArgProSerSerAr 2412
Db 7146 CACACTCACCTCGCCAGGTGGCGCGGGAAGGCCAAAGTCTCTGGCAGACCAGCAGCGG 7205
Qy 2412 GlySaAlaLysSerProAlaProGlyLeuAlaSerGlyYAspArgProProSerAlaSerSe 2432
Db 7206 AAAAGCAAGTCCCGGAGCCCGGCTGGCATGTGGAGACGGGCAACCTCTGTCTCTC 7265
Qy 2432 rValHlSerSerGluGlyAspCyAsnaRgaRgThrProLeuThraAspRgValTrpGluAs 2452
Db 7266 AGTGCACCTGGAGAGAGACTGCACACCGCGGACCGGCTCAACACCGGCTGGAGAG 7325
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyraSnProleuIlleMeArgLeuG 2472
Db 7326 CAGGCGCTGCTCGCAGGTGCCAGGCCATTCCTCCCTCAACCCCTCATCATGCGGTGCA 7385
Qy 2472 nAlaGlyValMeAlaSerProProProProGlyLeuProAlaGlySerGlyProleuAl 2492
Db 7386 GGGGGGTGTCATGTGCTTCCACCCCAACCGGCGCTTCCCGCGGAGAGCGGCGCTCGC 7445
Qy 2492 aglyProHlSaIaTrpAspGluGluProLysProleuLeuCySerGlnTyrgLuth 2512
Db 7446 TGGCCCCCAACGAGCTGGGAGCAGAGAGCCCAAGCCACTGTCTGTGCTCGAGTACAGAG 7505
Qy 2512 rLeuSerAspSerGlu 2517
Db 7506 ACTCTCCGACAGCGAG 7521

RESULT 6
US-09-819-104A-1
Sequence 1, Application US/09819104A
Publication No. US20030027137A1
GENERAL INFORMATION:
APPLICANT: Chen, J. Don
TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
TITLE OF INVENTION: AND USES THEREFOR
FILE REFERENCE: UMG-030
CURRENT APPLICATION NUMBER: US/09/819,104A
CURRENT FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: 60/193,138
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8686
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (157)..(7677)
US-09-819-104A-1

Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487

Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 10 Gaps: 3
US-09-522-753-5 (1-2517) x US-09-819-104A-1 (1-8686)

Qy 1 MetSerGlySerThrglnLeuValAlaGlnThrTrpArgAlaThrgluProArgTyPro 20
Db 157 ATGTGCGGCTCCACACAGCCTGTGGCAGAGGTGGAGGCGCACTGAGCCGCTTACCCG 216
Qy 21 ProHlSerLeuSerThrProValGlnIleAlaArgThrHlSaThraSPValGlyLeuLeu 40
Db 217 CCCCAAGGCTTCTTACCCAGTGCAGATGCGCCGAGCACAACGACGTCGAGCTCTCTG 276
Qy 41 GluTyrglnHlSaIaSerArgAspTyraLaseHlSaLeuSerProGlySerIleIleGln 60
Db 277 GAGTACAGACACACATCCCGGAGCTATGTCTCCACCTGTCCCGCGGCTCCATCATCCAG 336
Qy 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyYAsnGluArgSerGln 80
Db 337 CCCCAAGGCGGAGGCGCTCCCTGCTGTGAGTTCAGCCCGGGAATGAACGCTCCAG 396
Qy 81 GluLeuHlSaLeuArgProGluSerHlSaSerTyraLeuProGluLeuGlyLysSerGluMet 100
Db 397 GAGCTTCACCTCGGCGCAAGTCCCATATACCTGCCCCGAGCTGGGAAATCAGAGATG 456
Qy 101 GluPheIleGlySerLysArgProArgLeuGluLeuLeuProAspProleuLeuArgPro 120
Db 457 GAGTTCATTGAAGAAGAGCGCCCTGGCTAAGAGCTGCTCTGACCCCTGTGCGACCG 516
Qy 121 SerProleuLeuAlaThrglyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 517 TCACCCCTGCTGGCCACCGGCGAGCTGGGGATCTGAAGACCTCAACGAAGACCTTAC 576
Qy 141 LeuThrglyLysLeuGluProValSerProProSerProProHlSaThraSPProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAACCGGAGTCTCCCCAGCCCGCCGCACTACCTCTGAGCTG 636
Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMeAspArgValAsp 180
Db 637 GAGCTGTGTCGCCCAAGGCTGTCCAAGAGAGCTGATCAAGAACATGGAACCGCGTGAG 696
Qy 181 ArgGluIleThrMeValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
Db 697 CGAGAGATCACCATGTAGTACGACGATCTTAAGCTCAAGAAAGACGAAACGCTG 756
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
Db 757 GAGAGAGAGCTGCGCAAGCGCCGAGGCTGAGAAAGCCGCTGACCCGCGCCATCGAG 816
Qy 221 SerLysHlSaIaSerLeuValGlnIleIleTyraSPGluAsnArgLysLysAlaGluAla 240
Db 817 TCGAAGCACCGGAGCTGTGCAGATCATCTAGACGAACACCGAAGAAAGCTGAAGCT 876
Qy 241 AlaHlSaArgIleLeuGluGlyLeuGlyProGlnValGluLeuProleuTyraSnGlnPro 260
Db 877 GCACATCGGATTCGGAAGGCTGGGGGCCCGAGGTGAGACTCCGCTGTACAAACGAGCCC 936
Qy 261 SerAspThrArgGlnTyraHlSaGluAsnIleLysIleAsnGlnAlaMeArgLysLysLeu 280
Db 937 TCCGACACCCGGGAGATCATAGAACATCAAAATTAACAGAGCGATGCGAAGAGCTA 996
Qy 281 IleLeuTyraPheLysArgArgAsnHlSaIaArgLysGlnTrpLysGlnLysPheCysGln 300
Db 997 ATCTTGTACTTCAAGAGAGGAATCACGCTCGGAACAATGGAGAGCAAGATTCTCCAG 1056
Qy 301 ArgTyraSPGlnLeuMeGluAlaLeuGluLysLysValGluArgIleGluLysAsnPro 320
Db 1057 CGCTATGACCACTCATGAGGCTGGGAGAAAGAGTGAAGCGCATTCAGAAACAACCCC 1116
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyraGluLysGlnPheProGluIle 340

[illegible]

QY	701	GIuApbGIuGIuMeGIuAlaSerGIyAlaSerGIyAenGIuGIuMeValGIuGIu	720
Db	2254	GAGGATGAGAGATGAGAGCGCTGGCGCTGACGGGAATATGAGAGATGTGTGAGAG	2313
QY	721	AlaGIuAlaLeuHlaAlaSerGIyAenGIuValProArGIyGIuCySerGIyProAla	740
Db	2314	GCTGAA-----GCC	2322
QY	741	ThrValaAenSerSerAspThrGIuSerIleProSerProHlaThrGIuAlaAlaLys	760
Db	2323	ACTGTCAACCAACAGCTCAGACACCGAGAGCATTCCTCTCTCTCACACTAGAGGCCCAAG	2385
QY	761	AspThrGIyGIuAenGIyProLysProProAlaThrIleuGIyAlaAapGIyProProPro	780
Db	2383	GACACAGGGGACGAATGGGCCCAAGCCCCACACCTCGGGCGCGGACGGGCGACCCCCCA	2442
QY	781	GIyProProThrProPro-ArgArGIyThrSerArgAlaProIleGIuProThrProAlaSe	800
Db	2443	GGGGCACCCACCCACCAACCGGAGGACATCCGGGC-CCCACTAGTCCACCCCGGCTC	2501
QY	800	rGIuAlaThrGIyAlaProThrProProProAlaProProSerProSerAlaProProPr	820
Db	2502	TGAAGCCACTTAAAGCCCTTACGCCCCCAACGACCCCACTTCTCTCTCACTCTCTCC	2561
QY	820	QValaValProLysGIuGIuLysGIuGIuGIuThrAlaAlaAlaProProValaGIuGIuGI	840
Db	2552	TGTGTGTCCTCCACAGAGAGAGAGAGAGAGAGAGACCGCACACCGCCCTCAGTGGAGAGAG	2621
QY	840	YGIuGIuGIuLysProProAlaAlaGIuGIuLeuAlaValaAspThrGIyLysAlaGIuGI	860
Db	2622	GGAGGAGCAGAAAGCCCCCGCGCTGAGAGACTGGCAGTGAACACAGGAAAGCCGAGGA	2681
QY	860	uProValLysSerGIyCyStrGIuGIuAlaGIuGIuGIyProAlaLysGIyLysAspAl	880
Db	2682	GCCGTCACAAAGCCAGGTGCAGGAGAGAAAGCCGAGAGAGGGCGCCGACAAAGGCAAGAGC	2741
QY	880	AGIuAlaAlaGIuAlaThrAlaGIuGIyAlaLeuLysAlaGIuLysLysGIuGIyLysE	900
Db	2742	GGAGGCGGCTGAGGCGCACGCGCGAGAGGGCGCTCAAGGCAGAGAAAGAGAGGGCGGAG	2801
QY	900	rGIyAlaGIyAlaThrThrAlaLysSerSerGIyAlaProGIuAenSerAspSerSerAlaTh	920
Db	2802	CGGACAGGGCCACCAAGCCAAAGACTCGGGCGCCCCCAAGACAGCGACTTCAAGTCCAC	2861
QY	920	rCySerSerAlaAspGIuValaAspGIuAlaGIuGIyGIyAspLysAspArGIyLeuLeuSerPr	940
Db	2862	CTGCAGTGCAGACCAAGGTGATGAGCGCCGAGGGCGGCGACAAAGAACCGGCTGTCTCCC	2921
QY	940	QArpProSerLeuLeuThrProThrGIyAspProArGIyAlaAenAlaSerProGIuLysPr	960
Db	2922	AAGGCCCAAGCTCTCTCACCCGACTGGCGACCCCCCGGGCCAAATCCTCACCCCAAGAGCC	2981
QY	960	oleuAspLeuLysGIuLysGIuLysGIuAlaAlaAlaAlaIleProProIleGIuValaThrly	980
Db	2982	ACTGGACTTGAAGCAGCTGAAGCGACGAGGGCGGTCCATTCCTCCCACTCAGGTACCA	3041
QY	980	sValHlaGIuProProArGIuAspAlaAlaProThrLysProAlaProProAlaProPr	1000
Db	3042	AGTCATGATAGCCCCCGGGAGGAGCGAGCTCCACCAAGCAGACTCCCCAGCCCCAC	3101
QY	1000	oProProGIuAenLeuGIuProGIuSerAspAlaProGIuGIuProGIySerSerProAr	1020
Db	3102	GCCACCCCAAAACCTTGACGCGGAGACGACGCCCCCTCAGCAGCTGGCAGACACCCCG	3161
QY	1020	gGIyLysSerArGIySerProAlaProProAlaAapLysGIu-----	1033
Db	3162	GGCGCAAGACGAGAGCCCGGAGACCCCGCGCGACAAAGAGAGAGAGAAAGCTGTGTTCTT	3221
QY	1034	-----AlaPheAlaAlaGIuAlaGIuLysLeuProGIyAspProProCyStrThrSerGI	1052
Db	3222	CCGAGCCTTGGCAGACCGAGGCGCAAGAGCTGCTGGGGAGACCCCTTTCGTGAGATTCTCG	3281

QY	1052	YLeuProPheProValProProValGluValIleIleValAspProHisAlaProAspPr	1072
Db	3282	CCGCGCCCTTCCCCCGGCCCCCGGAGGTATCAAGGCTCTCCCGCATGCCGCGACCC	3341
QY	1072	oSer1aPheSerTYrAlaProProGluVH1AspLeuProLeuGluLeuHisAspTrAl	1092
Db	3342	CTCAGCCTTCTCTACGCTTCCACCTGGTCAACCTGCTGCCCTTG6GCTCCATGACATGC	3401
QY	1092	AArgProValLeuProArgProProThrIleSerAsnProProLeuIleSerSerAl	1112
Db	3402	CGGCGCGCTCTGGCGCGCCCAACCATCTCCAAACCGGCTCCCTCATCTCTCTGCG	3461
QY	1112	AlaHisAProSerValLeuGluValGlnIleGluValIleSerGlnGluMetSerValGl	1132
Db	3462	CAAGCACCCCAAGCGCTCTCGAAGGCAAAATAGTGCCATCTCCCAAGAAATGTGGTCCA	3521
QY	1132	nLeuHisValProTYrSerGluHisAlaIleValAProValGluProValThrmMetGlyLe	1152
Db	3522	GCTCCACGTCCCGTACTACAGACATGCCAAGGCCCGCGGTGG6GCTGTCAACATGG6GCT	3581
QY	1152	nProLeuProMetAspProTyrLysLeuAlaProPheSerGluValLysGlnGlnLe	1172
Db	3582	GCCCCGCGCATGACACCCCAAAAGCTGCACCTTCAGCGAGTGAAGCAGAGCAGCT	3641
QY	1172	uSerProArgGluGlnAlaGlyProProGluSerLeuGluValProThrAlaGlnGluAl	1192
Db	3642	GTCCCCACGGGGCCAGCGCTGGGGCCACCGAGAGCTGGGGGTGCCACAGCCAGGAGGC	3701
QY	1192	aSerValLeuArgGlyThrAlaLeuGluSerValProGluGlySerIleThrLysGlyI	1212
Db	3702	GTCCGTCTGAGAGGGAACAGCTCTGGGCTCAGTTCCGGGCGGAGACATCACAAAGCAT	3761
QY	1212	eProSerThrArgValProSerAspSerAlaIleThrTYrArgGlySerIleThrHisGl	1232
Db	3762	TCCACGACACAGGGGCGCCCTGGGACAGCGCATCAATACCGGGCTCATACCAACAGG	3821
QY	1232	YThrProAlaAspValLeuTYrLysGlyThrIleThrArgIleIleGlyGluAspSerP	1252
Db	3822	CACGCGACGTGACGTCTGTACAAAGGGAACCATCACAGATCATCTGGCGAGGACAGGCC	3881
QY	1252	oSerArgLeuAspArgGluYArgGluuAspSerLeuProLysGluVH1IleTYrGluGl	1272
Db	3882	GAGTGGCTTGACCCCGCGCGGAGGAGACAGCTGCCCAAGGGCCACGTCATCTACAAAGG	3941
QY	1272	YLysLysGluVH1sValIleuSerTYrGluGluVglMetSerValThrgIncySerLysGlu	1292
Db	3942	CAAGAAAGGCGCACGTCTGTCTTAAGAGGTGGCATGTCTGACACCAAGTCTCCAAAGA	4001
QY	1292	uAspGluYArgSerSerSerGlyProProHisGluThrAlaAlaProLysValGlnTYrAs	1312
Db	4002	GGACCGCGAAGACACACTCAGGACCCGCCCATATGACGGCGGCCCAAGCCACACTTANGA	4061
QY	1312	pMetMetGluGluYArgValGluYArgAlaIleSerSerAlaSerIleGluGluYLeuMetG	1332
Db	4062	CATGATGAGAGGGCCCGGTGGGAGGACCATCTCTCCACGACAGCATCGAAGGCTCATAGGG	4121
QY	1332	YArgAlaIleProProGluYArgHisSerProHisIleLeuYsgGlnGlnHisIleAr	1352
Db	4122	CCGTGCCATCCGCGCGGAGCGACACAGCCCCCACCACCTCAAAAGGACGACACATCCG	4181
QY	1352	gGlySerIleThrgInGlyIleProArgSerTYrValGluAlaGlnGluAspTYrLeuAr	1372
Db	4182	CGGGTCTCATCACAAAGGATCCCTCGGCTTACGTGAGGACAGAGAGACTACCTCGG	4241
QY	1372	GArgGluAlaLysLeuLeuLysArgGluGluTYrProProProProProProSerArgAs	1392
Db	4242	TGGGAGGCGCAAGCTCTTAAGACGGGAGGCGACGGCTTCGCGCCCAACGCTCAAGGGA	4301
QY	1392	pLeuThrgLualTYrLysThGlnAlaLeuGluProLeuYsgLeuLysProAlaHisGl	1412
Db	4302	CCTGACCGAGGCTCTACAGACCGCAAGGCTTG6GCCCCCTGAAAGCTGAAAGCGGCGCATGA	4361
QY	1412	uGluLeuValAlaThrValLysGluAlaGluYArgSerIleHisGluIleProArgGluGl	1432

Dd	4362	GGGCTTGGTGGCCACGGTGAAGAGGGGGCGCTCCATCATGAGTATCCCGCGAGGA	4421
Qy	1432	uLeuAArgHisThpProGluLeuProLeuAlaProArgProLeuIysGluIysSerIleTh	1452
Dd	4422	GCTCGGACACAGCCCGAGCTGCCCTTGGCCCCCGGCGCGCTCAAGAGAGGGCTCCATCAC	4481
Qy	1452	rGlnGlyThpProLeuIysTyTAspThrGlyAlaSerThThrGlyIysSerIlySHIshAS	1472
Dd	4482	GCAGAGGACCCCGTCAAGTACGACACCGGGGGCTCCACCATGTGCTCCAAAAGACAGCA	4541
Qy	1472	pValArgSerLeuIleGlySerProGlyArgThPhaProProValHisProLeuAspVa	1492
Dd	4542	CGTACGCTCTCTCATTCGGAGACCCCGCGCGAGCTTCCACCCGTGCACCCGCTGAGTGT	4601
Qy	1492	IMetAlaAspAlaArgAlaLeuGluArgAlaCyTyTGlUGluSerLeuIysSerArhPr	1512
Dd	4602	GATGGCCGACGCGCGGGACCTGAAGACGTGCTCTGACAGAGAGAGCTGAAGACCCGGCC	4661
Qy	1512	oGlyThAlaSerSerSerGlyIysSerIleAlaArgGlyAlaProValIleValProGl	1532
Dd	4662	AGGACCGCCACGAGCTCTGGGGGGCTCCATTGGCGGGCGGCCCGGGTCAATTGGCTGA	4721
Qy	1532	uLeuGlyAspProArgInSerProLeuThrTyGluAspHisGlyAlaProPheAlaGly	1552
Dd	4722	GCTGGGTAGCGCGGGCGAGAGCCCCCTTGACCTTAGAGGACACGGGGGACCTTTGGCGG	4781
Qy	1552	yHisLeuProArgIlySerProValThMetArgGluProThProArgLeuGlnGlu	1572
Dd	4782	CCACCTCCCAAGAGTTTCGCCCGTACCATTCGGGAGCCCAACGCGCGCTCGAGAGAGG	4841
Qy	1572	ySerLeuSerSerSerTyAlaSerGluAspArgIlyLeuThSerThProArgGluI	1592
Dd	4842	CAGCCTTTCGTCCAGCAAGGATCCCAAGACCGAAGACTGACGTGCAGAGCTTCGTGAGAT	4901
Qy	1592	eAlaIysSerProHisSerThValProGluHisHisProHisProIleSerProTyGly	1612
Dd	4902	CGCCAGTTCCTCCGACAGCACCGTGGCCGAGACACACCACTCCATCTGGCCTTAGA	4961
Qy	1612	uHisLeuLeuArgIlyAlaSerGlyValaIspLeuTyTArgSerHisIleProLeuAlaPh	1632
Dd	4962	GCACCTGTTTGGGGCGTGAATGGCGTGGACCTGATGACAGCACATCCCTTGCGCTT	5021
Qy	1632	eAspProThSerIleProArgIlyIleProLeuAspAlaAlaAlaAlaTyTyTLeuPr	1652
Dd	5022	CGACCCCACTTCATACCCCGCGCATCTCTGTGACGACACCGCTGCTCATACCTGCC	5081
Qy	1652	oArgHisLeuAlaProAspProThTyTProHisLeuIysTyTProTyTLeuIleArgGly	1672
Dd	5082	CCGAACTTGGCCCCCAACCCCACTTACCGGACCTGTACCCACCTTACTCATTCGGGG	5141
Qy	1672	TYTyTProAspThAlaAlaLeuGluAspArgGlnThIleIleAspTyTyleThrSe	1692
Dd	5142	CTACCCCGACACGGGGCGGCTGGAGAACCGGACAGCATCATATGACTCATATCACTTC	5201
Qy	1692	rGlnGlnMetHisHisAspThAlaThAlaMetAlaGlnArgAlaAspMetLeuArgI	1712
Dd	5202	GCAGAGATGACACCAACACGAGCGCACCGCATGGCCACGAGACTGATATGCTGAGGGG	5261
Qy	1712	yLeuSerProArgIlySerSerSerLeuAlaLeuAspThAlaAlaGlyProArgGlyIleI	1732
Dd	5262	CCTCTCGCCCCGAGGCTCTGTGCTGGACCTAACTACGCTGGGGTCCCGAGGGCATAT	5321
Qy	1732	eAspLeuSerGlnValProHisLeuProValIleuValaProProThPrProGlyThProAl	1752
Dd	5322	CGACTGTGCCAAGTGCACACACTGCTGTCTGTGTGCCCCGACACCAAGGACCCAGCG	5381
Qy	1752	aThrAlaMetAspArgLeuAlaTyTLeuProThAlaProGlnProPheSerSerArgHis	1772
Dd	5382	CACCGGCATGACCGCTTGTGCTACTTCCACACGCGCCCAAGCCCTTACAGACGCGCA	5441
Qy	1772	sSerSerSerProLeuSerProGlyGlyTyTProThHisLeuThIlyAsProThrThThSe	1792

Dh 5442 CAGACGCTCCCACTCTCCCGAGAGGTCCAAACACTTGAACAAACCAACACACGCTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluArg 1812
Db 5502 CTGGTCGAGACGGAGAGGAGACCGGGATCCAGAGCGGAGCCGGATCGGAGCGGAGAAA 5561
Qy 1812 sSerTLeuThrSerThrThrThrValGluHisAlaProTLeTPrArgProGlyThrG 1832
Db 5562 GTCCATCTCAGCGTCCACCAAGAGGTGAGAGCAGCAACCATCTGGAGACTGGTACAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCGAGAGAGCGGAGCGAGCGAGCAGCGCGGGGGTGGGGGAGCGAGCGCGCGCGCG 5681
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProTLeSerProArgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCACTCCCAATGCCAACAGCACTGGCCCATCTCCCTCGAGACCAAGATGCCCTCCA 5741
Qy 1872 nGlnArgProSerValLeuHisAsnThrGlyMetLeysGlyLeileThrAlaValGluPr 1892
Db 5742 GCAGAGACCCAGTGTGCTTCAACAACAGGATGAGATATCATCACCGCTGTGAGGCC 5801
Qy 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThr 1912
Db 5802 CAGCAAGCCACAGGATCTGAGGTCCACCTCACCTCTCAACCGTTGCCAGCTGCAC 5861
Qy 1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db 5862 ATTCCTCACTGCAACCACTGCCACTGGGGGAGCACTCGATGGAGGTCTAACCTTACCT 5921
Qy 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGAGAGCCGCTGTGCTGCCAAGAGAGCCCCCGGCTGCCCGGCAAGGGGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGACATGCTTCTCTCGCAAGCCCCCGGCTGCCGGGTGAGGCCCGC 6041
Qy 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
Db 6042 CTCTCTCCCGAGAGGGGCTCGAGGCCCGGCCCTTAGTCTCTCTCTCTGAGCCAGCC 6101
Qy 1992 aThrTLeaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProP 2012
Db 6102 CACCATCGCCGACCCCTCGAAGAACTCGCACTCACACGCAAGCCGAGCCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPheSe 2032
Db 6162 GGGGCACTGCTCGGCTCGGACCGGACCGGGAAMAGCTCAAGTAAACCTTTTTC 6221
Qy 2032 rTLeGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 6222 CATCCAGAACTGGAACTCCGTTCTTGAGTTACACGGCAGAGACTACAGCCCGAAG 6281
Qy 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProL 2072
Db 6282 GGTGGAGCCCGTCAGCCCTGTGAGCTCACCACTGACCCAGACAGAGGGCTCCCAA 6341
Qy 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
Db 6342 GCACTCGAAGAGCTGCAAGAGCACTGAGGGGAGCTCGGCCCAAGAGCAGCAG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLysLeuArgProLeuProG 2112
Db 6402 CCCCCTGAAGCTTGGGGGGAGGCCGCCCACTCCCAACCTGGCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAAGCCCTGCTCAGACCGGCTGCTCCAGACCGCCCAAGGGGTCAAGAGTCAACAG 6521
Qy 2132 lValThrLeuAlaGlnHisHisLeuGluValLeileThrGlnAspTyrThrArgHisHis 2152
Db 6522 GGTCACTTGGCCAGACATCATGAGGTATCATACAGAGACTTACACCGGAGCACACCC 6581

Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 AAGAGACTCAGGCCACCTCTGCCGCCCTCTTACTCTCTCTCTCTCTCTCTCTCTCT 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
Db 6642 GTTCTTGAACTCCGCGCGCCCAACAGTGACTTCTTACTCTCTCTCTCTCTCTCTCT 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyLysValArgSerProGluProAsnLysThr 2212
Db 6702 CCGGCGCTGTGCTCTCCCAAGCGAAGGGGCAAGAGGTCTTCAAGCCAAACAGAGC 6761
Qy 2212 rSerValLeuGlyGlyGlyLysAspGlyLeileGluProValSerProProGluGlyMetTh 2232
Db 6762 GTCCGCTTGGGTGTGTGTAGAGAGGTATTGAACCTGTGTCTCCACCGAGGGCATTGAC 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GAGAGCGAGGACCTCCGGAGGTGCTGTATCCGCTGTATCCGGGATGGGAGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyLysAsnThrSerGlnProProAlaPhePh 2272
Db 6882 GAGGCCAGAGGATGGGCTCCAAAGTCTCCAGGCAACACAGCCAGCGCAGCTTCTT 6941
Qy 2272 eSerLysLeuThrGlySerAsnSerAlaMetValLysSerLysLysGlnLysLeuAsnL 2292
Db 6942 CAGCAAGTACCGAGAGCACTCCGCAATGTCTCAAGTCCAAAGAGCAAGATCAACAA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCAACACCGAATGAGCTGTAATCAATATCAGCAGCCTGGAGACGA 7061
Qy 2312 uLePheAsnMetProAlaLeileThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCATATATGCTCCGCAATCAACGAGCCCTTAATGACCTTAAGAGCCAGGGGT 7121
Qy 2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeileArgLysAlaLeuMetG 2352
Db 7122 GCGAGAACATCGACAGCAACCAATGGGGCTGGAGGCCATTATTAAGAGCACTCATGG 7181
Qy 2352 yLysTyrAspGlnTrrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGATGGGAAGAGTCCCGCGCTCAGCGCCCAATGCTTTTAACTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProLleThrAlaAlaAspGlyArgSerAspHis 2392
Db 7242 TGGCAGTGCAGGCTTCCGCTGTATGCTCATMAACCGCTGTGAGCGAGAGGTACCA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyLysValLysValLysSerGlyArgProSerSerAr 2412
Db 7302 CACACTCAGCTGCCAGGTGGCGGGGAGGAGCCAGAGGTCTTGGAGACCCAGACCG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAAGCCAAGTCCCGCGCCCGGCTGTGAGTCTGGGAGCCGCGCACCTCTCTCTCTC 7421
Qy 2432 rValHisSerGluGlyLysPcyAsnArgArgThrProLeuThrAsnArgValTrrGluAs 2452
Db 7422 AGTGCACTGGAGGGAGACTGCAACCGCGAGCGGCTGCACCAACCGGCTGTGGAGGA 7481
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLleMetArgLeuG 2472
Db 7482 CAGGCCCTGTCTCGAGGTTCACAGCAATTCCTTCAACCCCTTATCATATGCGGCTGCA 7541
Qy 2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GGGGGGTGTCTATGGCTTCCCAACCCCAACGGGCTCTCCCGGGGAGCGGGCCCTCGC 7601
Qy 2492 aGlyProHisHisAlaTrrAspGluGluProLysProLeuLeuCysSerGlnTrrGluLuh 2512
Db 7602 TGGCCCCCAACAGCTTGGGAGAGAGAGCCAAAGCACTGTCTGTCTGCAGTACGAGAC 7661

OY 2512 rleuSerapSerGlu 2517
DB 7662 ACTCTCGACAGCGAG 7677

RESULT 7

US-10-174-014-4
Sequence 4, Application US/10174014
Publication No. US20040005292A1
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freiler
APPLICANT: Kenneth W. Doble
TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
FILE REFERENCE: PTS-0012
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 73
SEQ ID NO 4
LENGTH: 8686
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (157)...(7680)
US-10-174-014-4

Alignment Scores:

Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 18 Gaps: 3

US-09-522-753-5 (1-2517) x US-10-174-014-4 (1-8686)

OY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPRArgAlaThrGluProArgTyrPro 20
DB 157 AGTGGGGCTCCACACAGCTGTGGACACAGCTGGAGGGCCACTGAGCCCGCTACCCG 216
OY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 217 CCCACAGGCTTTCCTACCAAGTCAGATCGCCGGACGACGACGCGGCTCTCTG 276
OY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 277 GAGTACCAACACACACTCCCGGCACTATGCTCCCACTGTCGCGCTCCATCATCCAG 336
OY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGluValAsnGluArgSerGln 80
DB 337 CCCACGGGGGAGGGCCCTCTGCTGTGAGTTCAGAGCCCGGAATGACGATCCAG 396
OY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGluLysSerGluMet 100
DB 397 GAGCTCCACTGCGGCGACAGTCCCACTATACCTGCCGAGCTGGGGAAGTCAGAGATG 456
OY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 457 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAAGAGCTCTGCTGACCCCTGCTGCACCG 516
OY 121 SerProLeuLeuAlaThrGlyGlnProAlaGluSerGluAspLeuThrLysAspArgSer 140
DB 517 TCACCCCTGTGGCCACGGGCGAGCTGCGGATCTGAAGACTTCAACAGACCGTACG 576
OY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 577 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCAGGCCCCCGCACACTGACCTGAGCTG 636
OY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 637 GAGCTGTGCGCCGACAGGCTGTCTCAAGAGAGAGCTGATCCAGAAACATGACCGCGTGAC 696

OY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 697 CGAGAGATCACCATGTGTAGACACAGACTCTTAAGCTGAAGAGAAACACACACAGCTG 756
OY 201 GluGluGluValAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 757 GAGGAGGAGGCTGCACAGCCGCCGAGCTGAGAAAGCCGGTGTACCCGCCCTTCCAG 816
OY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysAlaGluAla 240
DB 817 TCGAAGCACCCGACGCTGTGTCAGATCTTACAGACGAAACCGGAAGAAGGCTGAGCT 876
OY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 877 GCACATCGGATTTCTGAAGAGCTGTGGGCCCGCAGGGTGGAGTGCCTCTTACCAACAGCC 936
OY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 937 TCCGACACCCGCACTATCATGAGAACATCAAAATTAACAGGCGATCGGAAGAACTA 996
OY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGluLysPheCysGln 300
DB 997 ATCTTGTACTTCAAGAGAGAGATCACCTCGAAGAACATGGAGACAGAGATTCTGCCAG 1056
OY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnPro 320
DB 1057 CGCTATGACACAGCTCATGAGGCGCTGGGAGAAAGTGGAGCGCATCGAAGAACACCC 1116
OY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGluPheProGluIle 340
DB 1117 CGGCGCGCGGCGAAGAGAGCAAGGTGCGAGTACTACGAGAACAGATTCCCTGAGATC 1176
OY 341 ArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgValGlyGlnArgLysGly 360
DB 1177 CGCAAGCAGCGGAGCTCGAGGAGCGCATGCG--AGGTTGGCGAGCGGCGAGTGGG 1233
OY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspLysLeuSer 380
DB 1234 CTGTCCATGTCCGCCCGCCGCGCAGCAGCAGAGGTGTGAGATCATCATGCGCTCTTA 1293
OY 381 GluGlnGluValAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1294 GAGCAGAGAAACCTGAGAGAGAGATGGCAGCTGGCGGTGATCCGCCCATGCTGTAC 1353
OY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1354 GACGCTGACACAGACCGCATCAAGTTCAACATGAACAGGCTTATGGCCGACCCAG 1413
OY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPhe 440
DB 1414 AAGGTGTACAAAGACCGCAGAGTCAATGAACATGTGAGTGAAGACAGAGAAAGACCTTC 1473
OY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuArg 460
DB 1474 CGGAGAAAGTTCAATGAGATCCCAAGAACTTTGGCTGTATGTCATCTCTGAGAGAG 1533
OY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysAsnGluAsnTyrLys 480
DB 1534 AAGACAGTGGCTGAGTGGCTCTTACTTACTGACTCAAGAAAGATGAGAACTAAG 1593
OY 481 SerLeuValArgArgSerTyrArgArgArgLysLysSerGlnGlnGlnGlnGln 500
DB 1594 AGCTGTGTAGACGCGAGCTATGCGCCCGCCGACAGACACAGCAAGAACAGCGACG 1653
OY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1654 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
OY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
DB 1714 GATGAG 1773
OY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyLysAsnAspGlu 560

|||||
Db 1774 GACACAGAAAGACCTTCCAGAGAAAGACACACCTCAGGGAGAGCAACAGACGAG 1833
Oy 561 LysGluValAlaValAlaSerIysGluArgIleThrAlaIleuSerGlnGlyArgIleGly 580
Db 1834 AAGAGAGCTGTGGCTCCAAAGGCGCGCAAACTCCCAACAGCGAGGAAGACGCAAGGCG 1893
Oy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1894 CGCATCACCCGCTCAATGGCTAATGAGGCCAAGCGAGGAGGCGCATCACCCCCACAGAG 1953
Oy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
Db 1954 AGCCCCAGCTGGCTCCATGAGACTGATAGAGTTCTCGCTGACAGAAAGAAATG 2013
Oy 621 GluThrAlaIleIysIysGluIleuGluGluIleGlyArgMetProSerAlaIleAlaArgMet 640
Db 2014 GAAACACGCCAAGAAAGGCTCTCTGAAACGCGCGCAACTGGTGGCCATCGCCGAGATG 2073
Oy 641 ValGlySerIysThrValSerGlnCysIleAsnPheTyrPheAsnTyrIleIysArgGln 660
Db 2074 GTGGGCTCCAAAGACTGTGTGGCACTGTAAAGACTTCTAATTCACTCAAGAAAGGCGAG 2133
Oy 661 AsnLeuAspGluIleLeuGlnGlnIleIysIleuIysMetGluIysGluArgAsnAlaArg 680
Db 2134 AACCTCATGAGATCTTGACAGACACAGCTGAAGATGAGAGAGAGAGACGCGCG 2193
Oy 681 ArgIleIysIysIysValProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db 2194 AGGAAAGAAAGAAAGGCGCGCGCGCGCGCGAGAGAGCTGCAATTCGCGCGGTGTG 2253
Oy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2254 GAGGATAGAGAGATGAGAGGCTCGGGCTGACGGGAATGAGAGAGATGTGTGAGAGAG 2313
Oy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyIysCysSerGlyProAla 740
Db 2314 GCTGAA-----GCC 2322
Oy 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaIys 760
Db 2323 ACTGTCAACAACAGCTAGACACCGAGAGCATCCCTCTCTCACTGAGGCGCGCAAG 2382
Oy 761 AspThrGlyIleAsnGlyProIysProProAlaThrIleuGlyAlaAspGlyProProPro 780
Db 2383 GACACAGGGAGAAATGGGCCCAAGCCCCAGCCACCTTGCGCGCGAGCGGCCACCCCA 2442
Oy 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db 2443 GGGCCACCCACCCACCAACCGAGAGACATCCGGCC-CCCACTGAGTCCACCCCGGCTC 2501
Oy 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACTTATGCCCCCTAGCGCCCCACAGCACCCCACTTCCCTCTCTCTCTCC 2561
Oy 820 oValValProIysGluGluIleIysGluGluIleThrAlaAlaAlaProProValGluGlu 840
Db 2562 TGTGTCTCCCAAGAGAGAGAGAGAGAGACCGACAGCGGCCCACTGAGAGAGGG 2621
Oy 840 yGluGluGluIleProProAlaAlaGluGluLeuAlaValAspThrGlyIysAlaGlu 860
Db 2622 GGAGAGACAGAAACCCCGCGGCTGAGAGCTGGAGTGCACACAGGAGGCGCAGGA 2681
Oy 860 uProValIysSerGluCysThrGluGluAlaGluGluGlyProAlaIysGlyIysAspAl 880
Db 2682 GCGCGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCGCCAAAGGCGCAAGGCGC 2741
Oy 880 agluAlaAlaGluAlaThrAlaGluGlyAlaLeuAlaGluIleIysGlyIysGly 900
Db 2742 GAGAGGCGCTGAGAGCCAGCGCGAGAGGCGCTCAAGCGAGAGAGAGAGAGGCGGAG 2801
Oy 900 rGlyArgAlaThrThrAlaIysSerSerGlyAlaProGluAspSerAspSerAlaThr 920

Db 2802 CGGAGGAGCCACACAGCCAAAGCTCGGGGCGCCCCCAGAGACAGCATCTCACTGCCAC 2861
Oy 920 rCysSerAlaAspGluValAlaAspGluAlaGluGlyIleAspIleAsnAlaGluLeuSerPr 940
Db 2862 CTGCAGTGCAGACAGAGGTGAGTGAAGGCGAGGCGGAGCAAGAAACGGCTGTCTCC 2921
Oy 940 rArgProSerIleuLeuThrProThrGlyIAspProArgAlaAsnAlaSerProGlnIysPr 960
Db 2922 AAGGCCAGCTCTCTCAACCCGACTGGCGAGACCCCGGGGCCAATGCTCAACCCGAAAGCC 2981
Oy 960 oLeuAspLeuIysGluIleuIysGluIleArgAlaAlaAlaIleProProIleGluValThrIy 980
Db 2982 ACTGACCTTAAGAGAGCTGAAGACGAGGCGGCTGCATCCCCCATCCAGGTCAACAA 3041
Oy 980 sValHisGluProProArgGluAspAlaAlaProThrIysProAlaProProAlaProPr 1000
Db 3042 AGTCATAGACCCCCCGGAGAGACGAGCTCCACCAACACAGCTCCCCAGGCCAC 3101
Oy 1000 oProProGluAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
Db 3102 GCCACGCCAABAACCTGACGCCGAGAGCGACGCCCTCAAGCAGCTGGACAGCCCCCG 3161
Oy 1020 gGlyIysSerArgSerProAlaProProAlaAspIysGlu----- 1033
Db 3162 GGGCAAGAGCAGAGCGCCGCGCACCCCGCGCAAGAGAGAGAGAGCTGTCTT 3221
Oy 1034 -----AlaPheAlaGluValAlaGluIleuIysLeuProGlyIAspProProCysThrThrSerG 1052
Db 3222 CCAGGCTTGCAGCGAGCGCCAGAGAGCTGCTGGGAGACCCCTTGTGAGATTCCGG 3281
Oy 1052 yLeuProPheProValProProArgGluValIleIysAlaSerProHisAlaProAspPr 1072
Db 3282 CTGCGCTTCCCGTGCCTCCCGGTAGAGTGAATCAAGGCTTCCCGCATGCCCGGACCC 3341
Oy 1072 oSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyIleuHisAspThrAl 1092
Db 3342 CTCAGCTTCTCTTACGCTCACCTGATCCACCTGACCCACTGCGCTCGCATACATGC 3401
Oy 1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3402 CGGCGCGTCTCGCGGCCACACCATCTCAACCCGCTCTCTCTCTCTCTG 3461
Oy 1112 alyHisProSerValLeuGluIleGlnIleGlyAlaIleSerGlnIymetSerValG 1132
Db 3462 CAAGACCCACAGGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGTGCTCCA 3521
Oy 1132 nLeuHisValProTyrSerGluHisAlaIysAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTCCAGGTCCGTACTCAGAGCATGCCAAAGGCCCGGTGGGCTGTCTACATGGGGCT 3581
Oy 1152 uProLeuProMetAspProIlyIleIleuAlaProPheSerGlyValIlyGluGluGlnLe 1172
Db 3582 GCCCTGCGCATGAGACCCCAAAAGCTGGACCTTCAAGCGAGTGAAGAGAGACGCT 3641
Oy 1172 uSerProArgGlyGluAlaGlyProProGluIleuGlyValProThrAlaGlnGluAl 1192
Db 3642 GTCCCAACGGGGCAGAGCTGGGCCACCGAGAGCTGGGGGTGCCACAGCCCAAGAGGC 3701
Oy 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIysGlyI 1212
Db 3702 GTCGCTGTGAGAGGAGACAGCTGTGGCTCAAGTTCCGGGGGAAAGCATCCAAAGGCAT 3761
Oy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisG 1232
Db 3762 TCCACACACAGGAGTCCCTCGAGACGCGCATCAATACCGGGCTCCATCCACCGG 3821
Oy 1232 yThrProAlaAspValLeuIleTyrIysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CAGCGCAGCTGAGCTCTGTACAAAGGCAACATACCAAGATATCGGAGAGACAGCCC 3881
Oy 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProIlyGlyHisValIleTyrGluG 1272
Db 3882 GAGTCCGCTTGACCCGCGCGGAGAGACAGCTGCCCAAGGCGCAGTCATTCAGAAAG 3941

QY 1272 YLysylSerGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSerIleArg1 1292
DB 3942 CAAGAAAGGCGCAAGTCTTGTCTATGAGGAGGTGGCATGTCTGTGACCCAGAGCTCCAAAGAA 4001
QY 1292 uAspGlyValGSerSerSerGlyProProHnIsgIuThrAlaAlaProLysArgThrTyrAs 1312
DB 4002 GGACCGGCAGAACAGCTCAGAGACCCCCCATGAGAGGCGCGCCCAACGCGCACTATGA 4061
QY 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGluYleuMetG1 1332
DB 4062 CATGATGAGGGGCGCGTGGGAGAGCCATCTCTCCAGCCAGATCGAAGGCTCTCATGGG 4121
QY 1332 YArgAlaIleProProGluArgHisSerProHnIshIleuLysGluGlnHisIshIleArg 1352
DB 4122 CCGTGCATCCCGCGGAGCAGACAGCCCCCAACCTCAAGAGCAGACACCATCCG 4181
QY 1352 gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluuAspTyrIleuArg 1372
DB 4182 CGGATCCATCACAAGAGGATCCCTCGTCTTACGTGAGGCAAGAGAGACTACCTGCG 4241
QY 1372 gATGAlaAlaLysLysLeuLysArgGluGlyThrProProProProProSerArgAs 1392
DB 4242 TCGGAGGCGCAAGCTCTTAAAGGGAAGGAGCAGCTCCGCCCCCAACCGCTCACGGGA 4301
QY 1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLysLysProAlaHisG1 1412
DB 4302 CCGAACCGAGGCGCTTAAAGGAGCGAGGCGCTCGGAGCTCCGTAAGCTGAAGCGGCGCATGA 4361
QY 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG1 1432
DB 4362 GGGCGCTGGGGCGCACGCTGAAGAGGCGGCGCGCTCATTCATGAAGATCCCGGCAAGAA 4421
QY 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThr 1452
DB 4422 GCTGCGGCAAGCGCCGAGCTGCTGCTGCGCCCGCGCGCTCAAGAGGCTCCATCAC 4481
QY 1452 rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
DB 4482 GGAGGGCACCGCGCTCAAGTACAGACCGGCGCGTCCACCATCGTCTCCAAAAGCACGA 4541
QY 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVal 1492
DB 4542 CGTACCTCTCTCATGTGGAGCGCCCGCGGAGCTTCCACCGCTCACCGCTGGATGT 4601
QY 1492 lMetAlaAspAlaArgAlaLeuGluAlaArgAlaCysTyrGluGluSerLeuLysSerArgPr 1512
DB 4602 GATGCGCGAAGCGCCGAGCACTGGAAGCTGCTCTCGAGAGAGCTCTGAAGAGCGGCGC 4661
QY 1512 oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG1 1532
DB 4662 AAGGAGCGCGCAGAGCTCGGGGGGCTCCATTGGCGCGGCGCGCGTATTGTGCTGA 4721
QY 1532 uLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaG1 1552
DB 4722 GCTGGGTAAGCCGCGGAGAGCCCTGTGACTATGAGAACCAAGGGGCAACCTTTTCCGG 4781
QY 1552 yHisIleuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG1 1572
DB 4782 CCACTCCCAAGAGTTCGCGCCGTGAACATGCGGAGGCCACCGCGCTGTGAGAGGG 4841
QY 1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluI1 1592
DB 4842 CAGCCTTTTCGTCCAGAGAGCATCCAGAGACGAAAGCTGACGTGACGCTGTGAGAT 4901
QY 1592 eAlaLysSerProHisSerThrValProGluHisIshIshIshIshIshIshIshIshIshIsh 1612
DB 4902 CGCAAGTCCCGCGCAGACAGCGTGGCCGAGCACCAACCACTATCTCGGCTCATGA 4961
QY 1612 uHisIleuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisG1leProLeuAlaPr 1632
DB 4962 GCACTGTCTTCGGGGGCTGAGTGGCTGACCTGTATCGAGGACATATCCCTTGAGCTT 5021

QY 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrIleuPr 1652
DB 5022 CCACCCCACTCATATACCCCGGCGCATCCCTCTGAGCGAGCGCGCTCTACTACTGCGC 5081
QY 1652 oArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgG1 1672
DB 5082 CGAGACCTGGCCCGCAACCCCACTTACCGGACCTGTACCACTTACTTACTTACTTACTTACTT 5141
QY 1672 YTyrProAspThrAlaAlaLeuGluuAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
DB 5142 CTACCCCGACAGCGGGCGCTGAGAACCGGACGACATCATCATATGACTATCATCTTC 5201
QY 1692 rGlnGlnMetHisIshIshThrAlaThrAlaMetAlaGlnAlaGlnAlaAspMetLeuArg1 1712
DB 5202 GCAGAGATGCACCAACACAGCGCACCGCATGGCGCCGAGCGAGCTGATATCTGAGGGG 5261
QY 1712 YLeuSerProArgGlySerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleI1 1732
DB 5262 CTTCTGCGCCCGAGATCTTCCTGGCACTCAATACGTGCGGGTCCCGAGGATCAT 5321
QY 1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
DB 5322 CGACCTGTCCCAAGTGCACACCTGCTGTGCTCGGCCGACACAGGACCCAGCAG 5381
QY 1752 aThrAlaMetAspArgLeuAlaTyrIleuProThrAlaProGlnProPheSerSerArgHis 1772
DB 5382 CACCGCATGAGACCGCTTGTCTTACTTCCACCGCGCCCGCCCTTGACAGCGGCA 5441
QY 1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSe 1792
DB 5442 CAGCAGCTCCCACTCTCCCGAGAGTCCAAACACTTGAACAAACCAACCAACAGCTC 5501
QY 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu 1812
DB 5502 CTGCTCGAGCGGAGCGAGACCGGGATCGAAGAGCGGAGCCGGATCGGAGCGGAGAAA 5561
QY 1812 sSerIleLeuThrSerThrThrThrValGluHisAlaProIleTPrArgProGlyThrG1 1832
DB 5562 GTCCATCTCCACAGTCCACACAGCGTGGAGAGCAGCACCATCTGAGACCTGTGACAA 5621
QY 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAl 1852
DB 5622 GCAAGACCGGAGAGCGAGCGAGCAGCGCGGGGTGGGGGAGCAGACAGCCGCCCGC 5681
QY 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG1 1872
DB 5682 CTCCCACTCCAGTGCACACAGACTCGCCCATCTCCCTCGGAGCCAGAGATGCCCTCA 5741
QY 1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
DB 5742 GCAGAGACCAAGTGTGCTTCAACAACACAGGCAATGAAGGTATCATACCGCTGTGAGCC 5801
QY 1892 oSerLysProThrValLeuArgSerThrThrSerThrSerProValArgProAlaAlaThr 1912
DB 5802 CAGCAAGCCCAAGTCTTAGGTCCACTTCCACTCTTCAACCGTTCCGCGCAGCTGCCAC 5861
QY 1912 rPheProProAlaThrHisCysAspLeuGlyGlyThrLeuAspGlyValTyrProThrIle 1932
DB 5862 ATTTCCACTGCACCACTGCGCCACTGGCGGAGCACTCGATGGGGTCTTACCTTACCT 5921
QY 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProPr 1952
DB 5922 CATGGAGCCCGTCTTGCTCCCAAGAGGCGCCCGGAGTGGCGCGGCGCAGAGCGGCGCG 5981
QY 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
DB 5982 AGCAGACCGGCGATGCTTCTCTGCGCAAGCCCCAGCCCGCTCGGGCTGGAAGCCCGC 6041
QY 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
DB 6042 CTCCTCCCGCAGAGAGGCTCGGAGGCGCGCGCCCTAGTGCCTCTGCTCGGACAGC 6101
QY 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisIshIshIshIshIshIshIsh 2012

```
Db 6102 CACCATGCGCCGCAACCCCTGGAGAAAGACCTCGACACTCACACCCGAGCCGAGCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAlaSerProHiAaRgLuLyethrGlnSerLySProPheSe 2032
Db 6162 GGGCCACCTGACCTGGCTCTGGACCCGACCCGAGAAAGACTAAAGTAAACCTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuLeuRgSerLeuGlyTyRhiSgLySerSerTyRserProGluG 2052
Db 6222 CATCCAGGAAGTGAACCTCGCTTCTGGGTTACAGGGGAGGAGCTACAGCCCGGAGG 6281
Qy 2052 yValGluProValaSerProValaSerSerProSerLeuThrHiSApLySgLyLeuProLy 2072
Db 6282 GGTGGAGACCCGCTGACCTGTGAGCTCACACCAAGTCAACCAAGAGGGGCTCCCAA 6341
Qy 2072 SHIleuGluGluLeuLeuApLySserHiSleuGluGlyGluLeuLeuRgProLySgInProG 2092
Db 6342 GCACCTGGAAAGAGTCCAGCAAGAGCCACTGGAGGGGAGCTGGGCCCCAAGAGCCAGG 6401
Qy 2092 yProValaLyLeuGlyGlyGluAlaAlaHiSleuProHiSleuRgProLeuProGluSe 2112
Db 6402 CCCCGTGAACCTTGGCGGGAGGCGGCCACCTCCACCTGGGCGCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThraAlaProGlyValaLySgLyhiSgInaRgVa 2132
Db 6462 CCAAGCCCTGCTCCAGCCCGCTGCTCCAGACCGGCCCAAGGGGTCAAGGTCAACAGCGGT 6521
Qy 2132 lValThLeuAlaGlnHiSeriSleuGluValaIleThrGlnaRgTyRThraRgHiSPr 2152
Db 6522 GGTCAACCTGAGCCAGCAATCAGTAGGTATCATCACAGAGTCAACCCGAGCCACCC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyRserPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCACTCAGCCGCAACCTGCGCCGCCCTTACTCTTCTTCTGGGCGAGCTGCC 6641
Qy 2172 oValleuAspLeuAaRgRgProProSerAspLeuTyRleuProProRaphiSgLyAl 2192
Db 6642 CGTCTGTGACCTCGCGCCGACCCAGTGAACCTTACTCTTCTTCTTCTTCTTCTTCTTCT 6701
Qy 2192 aProAlaAaRgLySerProHiSserGlyGlyGlyLyVaLyRgSerProGluProLeuLySth 2212
Db 6702 CCGGCGCCGCTGGCTCCCGCCAGCAAGGGGGGAGAGGTCTTCAAGCCCAACCAAGAC 6761
Qy 2212 rSerValaLeuGlyGlyGlyGluAaRgLyIleGluProValaSerProProGluGlyMetTh 2232
Db 6762 GTGCGTCTTGGTGTGTGTGAAGAGCTATTGAACCTGTCTCCACCGAGGGGAGATGAC 6821
Qy 2232 rGluProGlyhiSserRgSerAlaValTyRProLeuLeuTyRArgApGlyGluGlnTh 2252
Db 6822 GGAGCCGAGGCACTCCCGAGGTGTGTGATCCGCTGTATCCGGGATGGGAGACAGAC 6881
Qy 2252 rGluProSerRgRgMerGlySerLySserProGlyAaRThSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCGAGAGATGGGCTTCAAGTCTCAGGCAACCAAGCCAGCCGCGACCTTCTT 6941
Qy 2272 eSerLySleuThrGlnSerSerAlaMetValLySserLySgInGlnIleAaLy 2292
Db 6942 CAGCAAGCTACCGAGAGCAATCCGCCATGTCTCAAGAGAGAGAGATCAACA 7001
Qy 2292 vLySleuAaRThhiSleuAaRgAaGluProGluTyRAsnIleSeriGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCAACCAAGGATGAGCTGAATATCAATATCAGCCAGCTGGAGCGA 7061
Qy 2312 uIlePheAaRMetProAlaIleThrGlyThrGlyLeuMetThTyRArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATATCCGCGCATCAACCGAAGAGGCTTATATACCTTATGAAGCCAGGGGT 7121
Qy 2332 lGlnGlnHiSeriSeriThraAaRgMerGlyLeuGluAlaIleIleAaRgLyAlaLeuMetG 2352
Db 7122 GAGGAACATGCGAGCAACATGAGGCTGGAGGCAATATTAAGAAAGCACTCATGG 7181
Qy 2352 yLySryAaRgGlnTrpGluGlnSerProProLeuSerAlaAaRAlaPheAaRProLeuA 2372
```

```
Db 7182 TAAATATGACAGTGGAGAGATGCCCGCTCAGGCCCAATGCTTTTAACCTCTGA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThraAlaAlaAspGlyAaRSerAaRPh 2392
Db 7242 TGGCAGTGCAGGCTTCCCTGCTGTATGCCATTAACCTGTGTGACGAGAGACTGACA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyAlaLySgLyValaLySeriYARgProSerSerA 2412
Db 7302 CACACTCACTCCGACAGTGGCGCGGAGAGGCAAGGTCTGTGGAGACCAAGCCAGCCG 7361
Qy 2412 gLyAlaLySserProAlaProGlyLeuAlaSerGlyAaRgAaRgProProSerValaSer 2432
Db 7362 AAAAGCAAGTCCCGGCGCCGCGGCGCTGTGGGAGCCGACCTCTGTCTCTC 7421
Qy 2432 rValhiSserGluGlyAaRgPheAaRgAaRgTrpProLeuThraAaRgValTrpGluA 2452
Db 7422 AGTGACTCGAGAGAGAGACTGCAACCGCGAGCGGCTTCAACCAACCGCTGTGGAGGA 7481
Qy 2452 PATgProSerSerAlaGlySerThraProPheProTyRAsnProLeuIleMerAaRgLeuG 2472
Db 7482 CAGGCCCTGCTCCGAGGTTCAGAGCATTCCTCTTAACCCCTGATCATGGGGCTGCA 7541
Qy 2472 nAlaGlyValaMerAlaSerProProProProGlyLeuProAlaGlySeriGlyProLeuAl 2492
Db 7542 GCGGGGTGTATGCTTCCCTCCACCCGCGGCTTCCCGCGGCGGCGGCGGCGGCTTGC 7601
Qy 2492 aGlyProHiSeriAlaTrpAaRgGluGluProLySProLeuLeuCysSerGlnTyRgIuTh 2512
Db 7602 TGGCCCCACACAGCTGGGAGCAAGAGCCAAAGCACTGTCTGTCTGCGAGTACAGAC 7661
Qy 2512 rLeuSerAspSerGlu 2517
Db 7662 ACTCTCGACAGCGAG 7677

RESULT 8
US-10-723-860-1739
; Sequence 1739, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1739
; LENGTH: 8686
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1739

Alignment Scores:
Pred. No.: 0 Length: 8666
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 21 Gaps: 3

US-09-522-753-5 (1-2517) x US-10-723-860-1739 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValaIleGlnThraAlaAaRgIleuProAaRgTyRPro 20
Db 157 ATTTCCGGCTCCACAGCTGTGTGAGCAAGACTGAAGGCGCACTAAGCCCGCTACCCG 216
Qy 21 ProhiSerLeuSerTyRProValGlnIleAlaAaRgThraAaRgValaGlyLeuLeu 40
```


Db 217 CCCACAGCCTTTCTTACCCAGTGCAGATGCCGGAACGACAGCAAGCTCGGCTCTG 276
Qy 41 GtUyTgLnhiisierargApTyraIsernihleuserProgliserleileegin 60
Db 277 GAGTACACAGACACTCCCGCATGCTCCCACTGTCCGCGGCTTCATCATCCAG 336
Qy 61 ProGlnArGrArGrProSerleuLeuserGluPhenGlnProGluYanGluYanGrGln 80
Db 337 CCCAGCGCGCGGCTCTCCCTGCTGTCTGAGTTCCAGCCCGGAGATGAACGCTCCAG 396
Qy 81 GtUleuhiAleuArGrProGluSerhiiserTyTleuProGluuenuGlyluserGluMet 100
Db 397 GAGCTCACCTGGCGGACAGAGTCCCACTACTCTCCGAGCTGGAGAGTCAAGAGAG 456
Qy 101 GluPhenileGluSerlyArGrProArGleuGluLeuLeuProArSPProleuLeuArGrPro 120
Db 457 GAGTTCATTGAAGCAAGCGCCCTCGGCTAGAGCTGTGCTGACCCCTGCTGCGACCG 516
Qy 121 SerProleuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrlyAspArGrSer 140
Db 517 TCACCCCTGTGGCCACGGGCCACGCTGGCGGATCTGAAGACCTCAACCAAGACCTAGC 576
Qy 141 LeuThrGlyLyLeuGluProValSerProProSerProProhiSerhArSPProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAACCGAGTGTCTCCCCGAGCCCGCCGACACTGACCTTGAGCTG 636
Qy 161 GluLeuValProProArGrLeuSerlySGluGluLeuileGlnAsnMetArSPArGrValArSP 180
Db 637 GAGCTGTGTCCCGCAAGCTGTCTCAAGAGAGACTATCCAGAACATGGAACCGCGTGGAG 696
Qy 181 ArGrGluileThrMetValGluGlnGlnileSerlyLeuLylylySGlnGlnGlnLeu 200
Db 697 CAGAGATCAACCATGGTAGAGCAGCAGATCTTAACTGAAGAAAGCAGCAACGCTG 756
Qy 201 GluGluGluAlaAlaLySPProGluProGluLySPProValSerProProileGlu 220
Db 757 GAGAGAGAGGCTGCCAAGCGCCGAGACCTGAGAAAGCCGTGACCGCGCCCATCGAG 816
Qy 221 SerlyshiArGrSerleuValGlnileIeTyArSPGluAsnArGrlylylyValGluAla 240
Db 817 TCGAAGCACCGCAGCTGGTGGCAGATCATCTACAGAGAACCGGAAGAAAGCTGAAGCT 876
Qy 241 AlaHiArGrileuGluGlyLeuGlyProGlnValGluLeuProLeuTyArGrGlnPro 260
Db 877 GCACATCGGATCTCGAAGAGCTCGGGGCCCAAGTGAAGCTGCGCTGTAAACAACGAGCC 936
Qy 261 SerArSPThrArGrGlnTyHisiGluuaniIeIyshiAsnGlnAlaMetArGrlylylyLeu 280
Db 937 TCGAAGACCGGCGAGTATCATGAAACATCAAAATAAACAGGCGATGCCGAAGAAAGCTA 996
Qy 281 ILeuTyThrPhelyArGrArGrAsnhiAlaArGrLySGlnTyRlySGlnlyPheCySGln 300
Db 997 ATCTTGTACTTCAAGAGAGGAAATCAAGCTCGGAAACAANTGGAGCAAGATTCTCCAG 1056
Qy 301 ArGrTyArSPGluLeuMetGluAlaLeuGluLylylyValGluArGluileGluAsnArPro 320
Db 1057 CGCTAAGACAGCTCATGAGGCTCGGAGAAAGAGTGAAGCGCATCGGAACAACCCC 1116
Qy 321 ArGrArGrArAlaLySGluSerlySValArGrGluTyTyRlySGlnlySGlnPheProGluile 340
Db 1117 CGCGCGCGGCGCAAGAGAGCAAGTTCGCGAGTACTACAGAAAGCAATTCCTGAGATC 1176
Qy 341 ArGrLySGlnArGrGluLeuGlnGluArGrMetGlnSerArGrValGlyGlnArGrLySGly 360
Db 1177 CGGAACAGCGGAGCTGACAGAGCCCATGCAAG---AGGTTGGCCACGCGGCACTGGG 1233
Qy 361 LeuSerMetSerAlaAlaArGrSerGluHisiGluValSerGluileIleAspGlyLeuSer 380
Db 1234 CTGTTCATGTCCGCCCGCAGCGAGCAACGAGTGTCAAGATCATTCGATGCTCTCA 1293
Qy 381 GluGlnGluAsnLeuGluLySGlnMetArGrGlnLeuAlaValIleProPomeTLeuTyR 400
Db 1294 GAGCAGAGAACTTGAGAAAGCAGATGCCCACTGGCTGTATCCCGCCCATGCTGTAC 1353

Qy 401 AspAlaAspGlnGlnArGrileLySPheIleAsnMetAsnGlyLeuMetAlaAspPomeT 420
Db 1354 GAGCTGACACACAGCGCATCAAGTTCAACAATGAACGGGCTTATGGCCCAACCCCAAG 1413
Qy 421 LySValTyTlyLyAspArGrGlnValMetAsnMetTPSerGluGlnGluLySGlnTyRhe 440
Db 1414 AAGGTGTACAAAGACCGCAGAGTCAAGAACATGTGGAGTGAACAGAGAGAGAACTTTC 1473
Qy 441 ArGrGluLySPheMetGlnhiSPProlyAsnPheGlyLeuileAlaSerPheLeuGluArGr 460
Db 1474 CGGAGAACTTATCATCAGATCCCAAGAACTTTGGCTGATCGCATCATCTCTGGAGAG 1533
Qy 461 TyThrValAlaGluCyValLeuTyTyTyTlyLeuThrLylylyAsnGluYanTyTlyS 480
Db 1534 AAGACAGTGGCTGAGTGCCTCTTACTACTACCTGACTAAAGAAATGAATGAACTAAGA 1593
Qy 481 SerLeuValArGrArGrSerTyArGrArGrGlyLylySerGlnGlnGlnGlnGln 500
Db 1594 AGCTGTGTGAGACGAGCTATGTGGCGCGCGGCAAGGCCAGCACAGCAACAGCAGCAG 1653
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArGrSerSerGlnGluLylyS 520
Db 1654 CAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCCGACAGCCAGAGAGAGAA 1713
Qy 521 AspGluLySGlnLySGluLySGluAlaGlyLySGluGluGluLySPProGluValGluAsn 540
Db 1714 GATGAGAAAGAGAAAGAAAGAGGCGGAGAAAGAGAGAAAGCCGAGGTGAGAAAC 1773
Qy 541 AspLySGluAspLeuLeuLySGluLySPThrArSPThrSerGlyGluAspAsnArGrly 560
Db 1774 GACAAAGAAAGACCTCTCAAGAGAAAGACAGACACTTCAGGGAGAGCAACAGCAG 1833
Qy 561 LySGluAlaValAlaSerlySGlyArGrlySPThrAlaAsnSerGlnGlyArGrlySGly 580
Db 1834 AAGAGAGCTGTGGCTCCAAAGCCGCAAAATGTGCCAAACAGCCAGGAGAGCAAGAAAGCC 1893
Qy 581 ArGluIleThrArGrSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1894 CCATCAACCCGCTCAATGCTATGAGGCCAAGAGAGAGGCCATCAACCCCGCAGAG 1953
Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArGrTyRlySGlnGluMet 620
Db 1954 AGCGCGGAGCTGCTCCATGAGAGCTGAATGAGATCTTCGCTGACAGAAAGAAATG 2013
Qy 621 GluThrAlaLySGlyLeuLeuGlnhiSGlyArGrAsnTyRlySerAlaIleAlaArGrMet 640
Db 2014 GAAACAGCCAAAGAAAGTCTCTGGAAACAGGCGCAACTGGTCCGCTCCCGGATG 2073
Qy 641 ValGlySerTyThrValSerGlnCylylyAsnPheTyRheAsnTyTlylylylyValGln 660
Db 2074 GTGGCTCCAAAGCTGTGTGCGAGTGTAAAGAACTTCACTTCAACTACAAAGAGGCGAG 2133
Qy 661 AsnLeuAspGluileuGlnGlnhiSlylylyLeuLySPMetGluLySGluArGrAsnAlaArGr 680
Db 2134 AACCTCGATGAATCTTGAGAGCAACAGCTGAAGATGAAGAGAGAGAAAGCGCGG 2193
Qy 681 ArGrLylylylylylylyAlaProAlaAlaAserGluGluAlaAlaPheProProValVal 700
Db 2194 AAGAGAAAGAAAGAAAGCGCGCGCGCGCGCAGCAGAGAGGTCATTCGCCCGGTG 2253
Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGlnGluGluMetValGluGlu 720
Db 2254 GAGGATGAGAGATGAGCGCTCGGCGCGTGAAGGAGAAAGAGAGAGTGTGTGAGAG 2313
Qy 721 AlaGluAlaLeuhiAlaSerGlyYanGluValProArGrGlyGluCySerSerGlyProAla 740
Db 2314 GCTGAA-----GCC 2322
Qy 741 ThrValAsnAsnSerSerArSPThrGluSerIleProSerProhiSerhGluAlaAlaLyS 760
Db 2323 ACTGTCAACAAGCTCAACAACCGAGAGCAATCCCTCTCTCAACACTGAGGCGCCAAAG 2382

Qy 761 AspThrGlyGlnAanglyProLyAspProAlaThrIleuGlyAlaAspGlyProProPro 780
Db 2383 GACACAGGGGAGAAATGGGGCCCAAGGCCCAAGCCCTGGGGCGGAGCGGGCCACCCCA 2442
Qy 781 GlyProProThrProPro-ArgArgThrSerThrAlaProIleGlyUProThrProAlaSe 800
Db 2443 GGGGACCCACCCCAAG 2501
Qy 800 rGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACCTTAAGCCCTTAAGCCCTTAAGCCACCCCACTTCCCTTCACTCTCC 2561
Qy 820 OValAlaProLyGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 840
Db 2562 TGTGTGCTCCCAAG 2621
Qy 840 YGluGluGlnIleuSerProProAlaAlaGluGluIleuAlaValAspThrGlyValAlaGlu 860
Db 2622 GAG 2681
Qy 860 UProValIleuSerGlyIleuThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 880
Db 2682 GCCCGTCAAG 2741
Qy 880 AGluAlaAlaGluAlaThrAlaGluGlyAlaIleuIleuAlaGluIleuIleuGlyGly 900
Db 2742 GAG 2801
Qy 900 rGlyAlaThrAlaThrAlaIleuSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Db 2802 CGGAG 2861
Qy 920 rCySerAlaAspGlyAlaAspGlyAlaGluGlyGlyAlaAspIleuIleuIleuIleuSer 940
Db 2862 CTGAG 2921
Qy 940 OArgProSerIleuIleuThrProThrGlyAlaAspProArgAlaAspIleuIleuIleu 960
Db 2922 AAGGCCAGAGCTCTCAAG 2981
Qy 960 OLeuAspIleuIleuGlyIleuIleuGlyIleuAlaAlaIleuProIleuIleuIleuIleu 980
Db 2982 ACTGAG 3041
Qy 980 sValIleuGlyUProProArgGlyAlaAspAlaAlaProThrIleuProAlaProAlaPro 1000
Db 3042 AGTCATGAG 3101
Qy 1000 OProProGlnAspIleuIleuProGlyIleuAspAlaProGlnIleuProGlySerSer 1020
Db 3102 GCCACCGCAAAAGCTGAG 3161
Qy 1020 gGlyIleuSerIleuSerProAlaProProAlaAspIleuIleuIleuIleuIleuIleu 1033
Db 3162 GGGGAG 3221
Qy 1034 -----AlaPheAlaAlaGluAlaGlnIleuIleuProGlyAlaAspProProCysTrpThr 1052
Db 3222 CCCAG 3281
Qy 1052 YLeuProPheProValProProArgGlyAlaIleuAlaIleuAlaIleuAlaProAsp 1072
Db 3282 CTGAG 3341
Qy 1072 OSerAlaPheSerIleuAlaProProGlyAlaIleuProIleuIleuIleuIleuIleu 1092
Db 3342 CTAG 3401
Qy 1092 aArgProValIleuProArgProProThrIleuSerAspProProProIleuIleuSer 1112
Db 3402 CGGAG 3461
Qy 1112 alyPheIleuSerIleuIleuGlyAlaIleuIleuIleuIleuIleuIleuIleuIleu 1132

Db 3462 CAAG 3521
Qy 1132 nLeuIleuValProIleuSerGlyIleuIleuAlaProValGlyProValIleuIleuIleu 1152
Db 3522 GCTCAG 3581
Qy 1152 UProLeuProMetAspProLyIleuAlaProPheSerGlyValIleuGlnGlnIleu 1172
Db 3582 GCCCTGAG 3641
Qy 1172 USerProArgGlyGlnAlaGlyProProGlyIleuIleuIleuIleuIleuIleuIleu 1192
Db 3642 GTCCAG 3701
Qy 1192 aSerValIleuArgGlyIleuAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 1212
Db 3702 GTCCGTGAG 3761
Qy 1212 eProSerThrArgValProSerAspSerAlaIleuThrIleuArgGlySerIleuIleuIleu 1232
Db 3762 TCCAG 3821
Qy 1232 YThrProAlaAspValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1252
Db 3822 CAGGAG 3881
Qy 1252 OSerThrIleuAspArgGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1272
Db 3882 GAGTCGCTGAG 3941
Qy 1272 YIleuIleuGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1292
Db 3942 CAAG 4001
Qy 1292 UAspGlyIleuSerSerSerGlyProProIleuIleuIleuIleuIleuIleuIleuIleu 1312
Db 4002 GAG 4061
Qy 1312 pMetIleuGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1332
Db 4062 CATGATGAG 4121
Qy 1332 YArgAlaIleuProProGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1352
Db 4122 CGGTGAG 4181
Qy 1352 gGlySerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1372
Db 4182 CGGTGAG 4241
Qy 1372 gArgGlyAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1392
Db 4242 TCGGAG 4301
Qy 1392 pLeuThrGlyAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1412
Db 4302 CTGAG 4361
Qy 1412 uGlyIleuValAlaThrValIleuGlyAlaIleuIleuIleuIleuIleuIleuIleu 1432
Db 4362 GGGGAG 4421
Qy 1432 ULeuArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1452
Db 4422 GGTGAG 4481
Qy 1452 rGlnGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1472
Db 4482 GAG 4541
Qy 1472 pValArgSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1492

Db 4542 CGTAGCCTCCCTCATCGGAGCCCGCGGACGTTCCACCCGTCACCCGTCGATGT 4601
Qy 1492 IMer1aAspA1aArgA1aLeuGluArgA1aCysTyrGluGluSerLeuYserArpR 1512
Db 4602 GATGGCCGAGCCCGGACCTGGAGAGTGTCTGCTGAGAGAGAGCTGAAGAGCCGGCC 4661
Qy 1512 og1YThrAlaSerSerSerGlyYser1Iea1aArgGlyAlaProVal1IleValProG1 1532
Db 4662 AAGGAGCCGACAGAGCTCGGGGGGCTCCATTGGCGCGGGGCCCGCATTTGTGCTGA 4721
Qy 1532 uLeuGlyLysProArgGluSerProLeuThrTyrGluAAspH1aGlyAlaProPhel1aG1 1552
Db 4722 GTTGGGTAAAGCCGCGCAGAGCCCTGTGACCTAATGAGACCAAGGGGCAACCTTGGCCG 4781
Qy 1552 YH1sLeuProArgGlyLysProVal1ThrEArGluPProThProArgLeuGluG1 1572
Db 4782 CCACTCCCAAGAGTTCGCGCTGTGACATGCGGGAGCCACCGCGCTGTGAGAGGG 4841
Qy 1572 YSerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu11 1592
Db 4842 CAGCCTTTGTCAGAGAGGCAATCCAGAGCCGAAAGCTGACGTCAGCGCTGTGAGAT 4901
Qy 1592 eAlaLysSerProH1aSerThrTyrAlaProGluH1aSH1sProH1aPro1IeSerProTyrG1 1612
Db 4902 CCGCAAGTCCCGCAGACAGACCGTGCCTGACAGACCAACCAACCCATCTCGCTTATGA 4961
Qy 1612 uH1sLeuLeuArgGlyYValSerGlyYValAspLeuTyrArgSerH1a1eProLeuAlaH 1632
Db 4962 GCACTGCTTCGGGGGCTGAGTGGCTGTGACCTGTATCCAGGCAATCCCTTGACCTT 5021
Qy 1632 eAspProThrSer1IeProArgGlyLysLeuAspAla1aAla1aTyrTyrLeuPr 1652
Db 5022 CGACCCACCTCATACCCCGCGGCAATCCCTGTGAGCGAGCGCTGACTTACCTGCGC 5081
Qy 1652 oArgH1sLeuAlaProAsnProThrTyrProH1sLeuTyrProProTyrLeu1IeArgG1 1672
Db 5082 CCGACACCTGTGGCCCCCAACCCCACTACCGGACCTGTACCCACTTACTTCGCGG 5141
Qy 1672 YTYrProAspThrAla1aLeuGluAAsnArgGlnThr1Ie1eAsnAspTyr1IeThrse 1692
Db 5142 CTACCCCGACCGCGCGCTGTGAGAACCGGCAAGCATCATATGATCATACCTC 5201
Qy 1692 rGlnGlnMeCh1aSH1aAsnThrAlaThrAlaMeValaGlnArgAlaAspMeLeuArgG1 1712
Db 5202 GGAGCAGATGACACACACACCGGCCACCGCATGGCCAGCGAGCATGTGTCGAGGG 5261
Qy 1712 YLeuSerProArgGluSerSerLeuAlaLeuAenTyrAlaAlaGlyProArgGly1Ie11 1732
Db 5262 CCTCTGCCCCCGGAGTCTCTGTGGACATCACTAGCTGCGGGTCCCGAGGCAATCAT 5321
Qy 1732 eAspLeuSerGlnValProH1sLeuProValLeuVal1ProProThProG1YThrProAl 1752
Db 5322 CGACCTGTCCCAAGTGCACACTGCTGTGCTGTGTCGTCGCCGACACACAGGACCCGAGC 5381
Qy 1752 aThrAlaMeLAspArgLeuAlaTyrLeuProThrAlaProGlnProPheserSerArgH1 1772
Db 5382 CACCGGCATGAGACCGCTTGCTTACCTCCCAACCGGCCCCAGCCCTTCAGAGCGGCCA 5441
Qy 1772 sSerSerSerProLeuSerProGlyGlyProThrH1sLeuThrLysProThrThrTrse 1792
Db 5442 CAGCAGCTCCCACTCTCTCCAGAGGTTCCAAACACTTGAACAAACCAACCAACAGTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluY 1812
Db 5502 CTGTCTCGAGCGGAGAGGAGATCCGGATCGAGAGCGGAGATCGGAGATCGGAGGAAA 5561
Qy 1812 sSer1IeLeuThrSerThrThrThrValGluH1aAlaPro1IeTyrArgProGlyThrG1 1832
Db 5562 GTCCATCTCTACGTCACCAAGAGCGGTGAGACAGCAACCATCTGAGAGACTGTGTACAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCAGAGCAGCGGAGCAGCGGAGCAGCGGCGGGGTGGGGGAGCAGCAGCGCGCCCGC 5681

Qy 1852 aSerH1sSerH1aAlaH1aGlnH1sSerPro1IeSerProArgThrGlnAspAlaLeuG1 1872
Db 5682 CTCCCACTCCCATGCCACACAGCATCGGCCCATCTCCCTCGGAGCCAGAGATGCCCTTCA 5741
Qy 1872 nGlnArgProSerValLeuH1aAsnThrGlyMeLysGly1Ie1IeThrAlaValGluPr 1892
Db 5742 GCAGAGACCCAGTGTGCTTCAACAACAGCAGCATGAAGGATATCATACCGCTGTGAGCC 5801
Qy 1892 oSer1YsProThVal1LeuArgSerThrSerThrSerSerProValArgProAla1aTh 1912
Db 5802 CAGCAAGCCCAAGTCTCTGAGTCCACTTCCACTCTCTACCCGTTCCGCGCAGCTGCAC 5861
Qy 1912 rPheProProAlaThrH1sCysProLeuGlyGlyThrLeuAspGlyYValTyrProThr1e 1932
Db 5862 ATTCACACTGCACCAACTGCCCCAAGTGGCGGACCCCTCGATGGGGTCTTACCTTACCT 5921
Qy 1932 uMeGluProVal1IleuLeuProLysGluAlaProArgValAlaAspProGluAAspProAr 1952
Db 5922 CATGAGCCCGTCTTGTGCTCCAGAGAGGCCCCCGGGTGCCTGGCCAGAGCGGCCCG 5981
Qy 1952 gAlaAspThrGlyH1aAlaPheLeuAlaLysAspProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCATGCTTCTCTCGCAAGCCCAAGCCCGCTCGGGCTGAGGCCCGC 6041
Qy 1972 aSerSerProSerLysGlySerGluProArgProLeuVal1ProProValSerGlyH1a1 1992
Db 6042 CTCCTCCCCACAAAGGCTCGGAGCCCGGCCCTAGTGCCTCTGTCTGTGCGCACGC 6101
Qy 1992 aThr1IeAlaArgThrProAlaLysAsnLeuAlaProH1aSH1aLaserProAspProPr 2012
Db 6102 CACCATGCGCCGACCCCTGTGCAGAGAACTCGACTTCCACACCGCCAGCCGAGCCCGC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProH1sArgGluYsrGlnSerLysProPhese 2032
Db 6162 GGCGGCACTGCTCGGCTCGGACCCGACCGGAGAAAGACTCAAAAGTAAACCTTTTC 6221
Qy 2032 r1IeGlnGluLeuGluLeuArgSerLeuGlyTyrH1aGlySerSerTyrSerProGluG1 2052
Db 6222 CATCCAGAACTGGAATCCGTTCTGTGGTTTACCAACGACAGCTTACAGCCCCCAAG 6281
Qy 2052 YValGluProValSerProValSerSerProSerLeuThrH1aAspLysGlyLeuProLy 2072
Db 6282 GGTGAGCCCGTCAAGCCCTGTGACTCACAGTGTGACCCACAGAAAGGGGTCTCCCA 6341
Qy 2072 SH1sLeuGluGluLeuAspLysSerH1sLeuGluGlyGluLeuArgProLysGlnProG1 2092
Db 6342 GCACTGGAAAGCTGCAAGAGCACTTGAAGGGGAGCTGGCCCAAGCAGCCAG 6401
Qy 2092 YProValLysLeuGlyGlyGluAlaAlaH1sLeuProH1sLeuArgProLeuProGluSe 2112
Db 6402 CCCCCTGAAGCTTGGGGGAGGCGGCCCACTCCACACTTGGCGCCCTGTCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyYValLysGlyH1aGlnArgVa 2132
Db 6462 CAGGCCCTGTCAAGCCCTGTCTCCAGACCGGCCAGGGGTCAAGATCACCAGCGGGT 6521
Qy 2132 1ValThrLeuAlaGlnH1a1eSerGluVal1IeThrGlnAspTyrThrArgH1aSH1aSP 2152
Db 6522 GGTCACTCGGCCAGCAATCATGAGTCTATCACAGACTACACCGGACCAACCC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyYAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAAGGCAACCCCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6641
Qy 2172 oValLeuAspLeuAArgArgProProSerArgLeuTyrLeuProProProAspH1aGlyAl 2192
Db 6642 CGTCTGAGACTTCGCGCCGACCAAGTACTTACCTTCGCGCCCGGACCATGCTGC 6701
Qy 2192 aProAlaArgGlySerProH1sSerGluGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCGGGCCGTGCTCCCCCAAGCAAGGAGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6761

QY 2212 rSerValleuGlYglYgluaPglYleGluproValSerProgluGlYmetTh 2232
Db 6762 GTCCGCTCTTGAGTGGTGTGAGACGAGTATGAACCTGTGCCCCACCGGAGGCGCATGAC 6821
QY 2232 rGluproGlYHISerTgSerAlaValTYrProleuLeuTYrTgAapglYgluGlInTh 2252
Db 6822 GGAGCCAGGGGACTCCCGGAGTGTGTGTACCCCTGCTGTACCGGAGTGGGAAACGAGC 6881
QY 2252 rGluproSerTgMetglYSerTgSerProglYASnThSerGlInProAlaPhePh 2272
Db 6882 GGAGCCAGGAGTATGGCTCTCAAGTCTCCAGGCAACCGACCGCGGCACTTCTT 6941
QY 2272 eSerTgLeuThrGlusSerASnSerAlaMetValYSerTgAlaPheGlInleuAnly 2292
Db 6942 CAGCAAGCTGACCGAGAACCTCCGCGATGTCAAGTCAAGAGAACGATCAACAA 7001
QY 2292 gLySleuLeuThrHISnnaTgAenGluproGlUTYrASnTleserGlInProglYThrGl 2312
Db 7002 GAAGCTGAACACCCACACCGGAATGAGCTGAAATACATATACGCGAGCTGGGAGCGGA 7061
QY 2312 uilePheASnMeProAlaIleThrGlYThrGlYleuMeThrTYrTgSerGlInAlaVa 2332
Db 7062 GATCTTCAATATGCCCGCATACCGAACAAGGCGCTTATGACTATGAAAGCGAGCGGT 7121
QY 2332 lGIngluHISAlaSerThrASnMeTgYleuGluaAlleIleArySAlaLeuMeTGl 2352
Db 7122 GCAAGAACATGCCAGCAACCAATGGGGCTGGAGCCATATTAAGAAAGCACTCATGGG 7181
QY 2352 yLYrTYrAapglInTggluGlusErProProleuSerAlaASnAlaPheASnProleuAS 2372
Db 7182 TAAATATGACACAGTGGGAAAGTCCCGCGCTCAGCGCCATCTTTAACTCTGAA 7241
QY 2372 nAlaSerAlaSerleuProAlaAlaMetProIleThrAlaAlaASnGlYrTgSerAapH 2392
Db 7242 TGCCAGTCCAGGCTGCGCTGTATGCCCCATTAACGCTGTGACGAGAGTGTGCA 7301
QY 2392 sThrLeuThrSerProglYglYglYlySAlaValSerGlYrTgProSerSerAr 2412
Db 7302 CACACTCACTCGCAGAGTGGCGCGGGAAGGCCAAGGTCTCTGCGAGACCCAGCACCG 7361
QY 2412 gLYrAlaLYSerProAlaProglYleuAlaSerGlYASpArpProProSerValSerSe 2432
Db 7362 AAAAGCCAAATCCCGCGCGCGGCTGCGATCGGAGACCGGACCTCTGTCTCTC 7421
QY 2432 rValHISerGlulYASpCYASnAArgThrProleuThrASnAryAlaTggluAS 2452
Db 7422 AGTGCATCGAGAGAGACTGCAACCGCGGACCGCTCAACCGCGTGTGGAGGA 7481
QY 2452 pATgProSerSerAlaGlYSerThrProPheProTYrASnProleuIleMeArGlueGl 2472
Db 7482 CAGGCGCTGTCGCGAGGTTCCAGCGCATTCCTCTACACCCCTGATCATGGGCTGCA 7541
QY 2472 nAlaGlYAlaMeAlaSerProProProProglYleuProAlaGlYSerGlYProleuAl 2492
Db 7542 GGGGGGTGATGAGCTTCCCGACCCCGCGGCTCCCGCGGGGACGGGCGCTCGC 7601
QY 2492 aGlYProHISAlaIleThrAapgluGluproLYSProleuLeuCYSerGlInTYrGluth 2512
Db 7602 TGGCCCCACACGCGCTGGGAGCGAGGCCAACGCACTGCTGTGCGAGTACGAGAC 7661
QY 2512 rLeuSerAapSerGlU 2517
Db 7662 ACTCTCGAGACGCGAG 7677

RESULT 9
US-10-723-860-6114/c
; Sequence 6114, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193 .NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent version 3.2
; SEQ ID NO 6114
; LENGTH: 9079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-6114
Alignment Scores:
Pred. No.: 0 Length: 9079
Score: 12845.50 Matches: 2473
Percent Similarity: 98.10% Conservative: 6
Best Local Similarity: 97.86% Mismatches: 20
Query Match: 97.20% Indels: 31
DB: 21 Gaps: 3
US-09-522-753-5 (1-2517) x US-10-723-860-6114 (1-9079)
QY 1 MetSerGlYSerThrGlInleuValAlaGlInThrTgPArGAlaThrGluproArGTYrPro 20
Db 8552 ATGTCGGGCTCCACACAGCGCTGTGGCACAAGTGGAGGGCCACTGAGCCCGCTTACCG 8493
QY 21 ProHISerleuSerTYrProValGlnIleAlaArgThrHISThrAapValGlYleuLeu 40
Db 8492 CCCACAGCTTTCCTACCAAGTGCAGATGCCCGGACGACACAGCGAGCTGGGCTCTCG 8433
QY 41 GlUTYrGlnHISerASnAArgTYrAlaSerHISleuSerProglYSerIleIleGlIn 60
Db 8432 GAGTACACAGACACACTCCCGACTGATGCTCCACCTGTGCGCGCTCATCATCAG 8373
QY 61 ProGlnAryAryAryProSerleuLeuSerGlUpheGlInProglYASnGlArySerGlIn 80
Db 8372 CCCAGGGGAGAGCGCTCTCGTGTCTGAGTTCAGCCCGGAAATGAACGCTCCAG 8313
QY 81 GlulLeuHISleuAryProgluSerHISerTYrleuProgluLeuGlYlySserGlumet 100
Db 8312 GAGCTCCACCTGGGCGACAGAGTCCCACTACCTGCCGAGCTGGGGAAGTCAGAGATG 8253
QY 101 GlupheIleGlusSerTgAryAryProAryleuGlInleuLeuProAryProleuLeuAryPro 120
Db 8252 GAGTTCATTGAAGCAAGCGCTCGCTGAGAGCTGTGCTTGAACCTCCAGCG 8193
QY 121 SerProleuLeuAlaThrGlYGlInProAlaGlYSerGlUASpLeuThrLYSAPArSAr 140
Db 8192 TCACCCCTGTGGCCACAGGCGCGCTGCGGATCTGAAGACTTCAACCAAGACCGGCC- 8134
QY 141 LeuThrGlYlySleuGluproValSer- ProProSerProProHISThrAapProglUe 160
Db 8133 CTGACGGGCAAGCTGGAACCGGTGTCTCCACCAACCCCGCACACTGACCTGAGCT 8074
QY 160 uGlulLeuValAProProAryleuSerTgYlySglulLeuIleGlnASnMeArpAryValAe 180
Db 8073 GAGGTGTGTCCCGCAGCGCTGTCCAAAGAGAGCTGATTCACAACCTGGAACCGCGTGA 8014
QY 180 pATgGlulIleThrMeValGlulGlnIleSerTgSleuLYSleuLYSleuGlnGlnIle 200
Db 8013 CCGAGAGATCACATGTGTAGAGACAGATCTTAAAGCTGAAGAAGACAGACAGACT 7954
QY 200 uGlulGlnIleAlaLYSProProgluproglYlySProValSerProProIleGl 220
Db 7953 GAGAGAGAGAGCTGCCAAGCGCGCGAGCTGGAACCGGTGTCAACCGCGCGCATCGA 7894
QY 220 uSerLYSHISerSleuValGlnIleIleTYrAapgluASnAryLYSAlaGlua 240
Db 7893 GTCGAAGACCGAGCGCTGTGTGAGATCATCTTACGACGAGAACCGGAAGAGCTGAAGC 7834
QY 240 aAlaHISarGlleuGlulYleuGlYProGlnValGlulLeuProleuTYrASnGlInPr 260

|||||
Db 7833 TTCACATCCGATTCTGGAAAGGCTGGGGCCCGCAGTGGAGCTGCCCTGTACCAACAGACC 7774
OY oSerAspThrArgIntYrHiIsgluasn1IelysIIeaangIaIaMetArgIylsIle 280
Db 7773 CTCGACACCCGGCATATCATGAGAACATCAAAATAAACAGGCAATGGCAAGAAAGCT 7714
OY uIleuIyYrPheIyAsArgAsnhiIsaIaArgIysgiIntPlysgIuIyPheCysGI 300
Db 7713 AATCTTGTACTTCAAGAGAGAAATCACCGCTCGGAACAATGGAGACAGAAAGTTCTGCCA 7654
OY 300 nArgIYrAspGIuIeMetGIuAlaIeugIuIyIsvIaGIuArgIIeGIuAsnAsnPr 320
Db 7653 GCGCTATGACCAAGCTCATGAGGCTCGGAGAAAGGTGGAGCCGATCGAAGAACACC 7594
OY 320 oArGArGAsGAlaIySGIuSerIySvAlaArgIuIYrGIuIySGIuPheProGIuII 340
Db 7593 CCGGCGGCGGGCCAGAGAGCAAGATTCGCGAGTCTACAGAGACAGTCCCTAGAT 7534
OY 340 eArgIySGIuArgIuIeugInGIuArgMeGIuSerArGIaGIyGIuArgGIySergI 360
Db 7533 CCGCAAGACGCGGAGCTGAGAGAGCGCATGAG---AGGTTGGCCAGCGGGGAGTGG 7477
OY 360 yIeuSerMetSerAlaIaArgSergIuHiIsgIuValSergIuIleIIeaSpGIyIeuSe 380
Db 7476 GCTGTCCATGTGCGCCGCGCAGCAGACGAGGTGTCAAGATCATGATGCTGCTTC 7417
OY 380 rGIuGIuGIuAsnIeugIuIySGIuMeArGIuIaIaValIIProProMetLeuY 400
Db 7416 AGAGCAGAGAACCTGAGAGACAGATGCCCGAGCTGGCGGTATCCCGCCCATGCTGTA 7357
OY 400 rAspAlaAspGIuGIuArgIIeIySphelIeaenMetAsnGIyIeuMeIIaAspProme 420
Db 7356 CGACCGTGAACAGCAGCGGATCACTTCAATCAATGAACGGGCTTATGGCCGACCCCAT 7297
OY 420 tIySvAlIYrIyAspArgGIuValMeIeMetTTPSerGIuGIuGIuIySGIuIhPrh 440
Db 7296 GAAGGTGTCAAAAGACCGCAGGTGTGAACATGTGAGTGAACAGAGAAAGAACCTT 7237
OY 440 eArgGIuIySphMeGIuHiIspProIySAsnPhelIyIeulIaIaSerPheIeugIuXr 460
Db 7236 CCGGAGAAAGTTATCATGACATCCCAAGAACTTTGGCTGTGATGCACTATTCTCGAGAG 7177
OY 460 gIySthIaIaIaGIuCySvAlIeutyYrYrIyIeuthIyIySAsnGIuAsnYrIy 480
Db 7176 GAAGACAGTGGCTGAGTGGCTCTTATTACTACCTGACTAAGAGATGAGAACTATTA 7117
OY 480 sSerIeuValaIrgArSergIYrArgArGArgIyIySergInGIuGIuGIuGIuGI 500
Db 7116 GAGCGGTGTGAGACGAGGCTATCGGCGCCGCGCAAGAAC--CAGCAGCAAACAAAAG 7059
OY 500 ngInGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 520
Db 7058 CCGGAGGCGACGCCACCAAGCCAGCCAGCCATGCCCCGCGACGCCAGAGAGAA 6599
OY 520 sAspGIuIySGIuIySGIuIySGIuIaGIuIySGIuGIuGIuGIuGIuGIuGIuGIu 540
Db 6598 AGATGAGAAAGAGAAAGAAAGAGGCGAGAAAGAGAGAGAAAGCCGAGGTGGAGAA 6539
OY 540 nAspIySGIuAspIeuleuIySGIuIySthIraAspAspThrSerGIyGIuAspAsnAspGI 560
Db 6538 CGACAAAGGAAGACTCTCTCAAGAGAAAGACAGACACTCAGGGGAGAGCAACAGCA 6879
OY 560 uIySGIuAlaValaIaSerIySGIyArgIySthIraIaAsnSergInGIyArgArgIySGI 580
Db 6878 GAAGGAGGCTGTGGCTTCAAAAGCGCAAACTGCCAACAGCAGGAGAAAGCAAGAAAG 6819
OY 580 yArgIIeIhIraSergMeIIaAsnGIuIaIaAsnSergIuGIuIaIleIhIraProGIuGI 600
Db 6818 CCGCATACCCGCTCATGTGCTATAGAGCCAAAGAGAGAGGCAATCACCCCCAGCA 6759
OY 600 nSerAlaGIuIeulIaSerMetGIuIeuaenGIuSerSerArGTTPThrGIuGIuGIu 620

Db 6758 GAGCCCGCAGCTGGCTTCATGAGAGCTGAATGAGATTCTCCCTGGACAGAAAGAAAT 6699
OY 620 rGIuThIraIaIySlySGIyIeuleuGIuHiIsgIyArGaenIrpSerAlaIaIaIaIy 640
Db 6698 GGAACAGCCCAAGAAAGGTCTCTCGSAAACAGCGCCGCAACTGTGGCCATCGCCCGAT 6639
OY 640 tValGIySeryIySthIraIaSerGIuCyIyAsnPhelYrPheIeIhIyIyIyAsArgGI 660
Db 6638 GGTGGCTTCAAGACTGTGTCCAGTGAAGAACTTCACTTCACTTCACTTCACTTCACTTCA 6579
OY 660 nAsnIeuaSpGIuIleuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 680
Db 6578 GAACCTCGATGATGATTTTCAGACAGACAAAGCTGAAGATGAGAGAGAGAGAACCCGG 6519
OY 680 gArGIuIySlyIySvAlaIaProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 700
Db 6518 GAGAAAGAAAGAAAGCGCCGCGCGCCAGCAGAGAGGTGATTCGCCGCCGTGGT 6459
OY 700 IGIuAspGIuGIuMetGIuAlaSerGIyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 720
Db 6458 GAGAGATGAGAGATGAGCGTCCGGCGTGAACGGAAATGAGAGAGATGTGGAGGA 6399
OY 720 uIaGIuAlaIeulIaIaSerGIyAsnGIuValProArGIyGIuCySergIyProAl 740
Db 6398 GCGTGA-----GC 6390
OY 740 aThIraIaAsnAsnSerSerAspThrGIuSerIIeProSerProhIeThrGIuAlaIaIy 760
Db 6389 CACTGTCAACAAAGTGTGAGACCGAGAGCATCCCTCTCTCTCACTAGAGGCGCCCA 6330
OY 760 sAspThrGIyGIuAsnGIyProIySProProAlaThIreugIyAlaAspGIyProProPr 780
Db 6329 GGAACAGGCGAGAAAGGGCCCAAGCCCAAGCCCACTGGGGCGGAGCGGCGACCCCT 6270
OY 780 oGIyProProPrProPro-ArgArgThSerArgAlaProIIeGIuProProPrAla 800
Db 6269 AGGCGCACCAACCCACACCGAGAGACATCCCGCC-CCCACTAGTCCACCCCGGCT 6211
OY 800 eGIuAlaIhThGIyAlaProThrProProProAlaProProSerProSerAlaProPr 820
Db 6210 CTGAAGCCACTTAAGCCCTTAAGCCCCCAAGACACCCCAATTCCTCTTCACTCTC 6151
OY 820 roValaIa 840
Db 6150 CTGTGTCTCCCAAG 6091
OY 840 IyGIuGIuGIuIySProProAlaIaIaGIuGIuIeulIaIaIaIaIaIaIaIaIaIaIaIa 860
Db 6090 GGAAG 6031
OY 860 IuProValIySserGIuCySthIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 880
Db 6030 AGCCGCTCAAGAGCGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5971
OY 880 IaGIuAlaIa 900
Db 5970 CGAGAGCCCTGAGGCGCAAG 5911
OY 900 eGIyArGIa 920
Db 5910 GCGGAGGCGCACAG 5851
OY 920 hIcYsSerAlaAspGIuValaAspGIuAlaGIuGIyGIyAspIySAsnArGIuIeuleuSer 940
Db 5850 CCGAGTCAAGAGCGAGTGAAG 5791
OY 940 roArpProSerIeuleuThrProThrGIyAspProArgAlaAsnAlaSerProGIuIyP 960
Db 5790 CAAGGCCAGCTCTTCAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5731
OY 960 roIeuaSpIeuleuIySGIuIySGIuIySGIuIaIaIaIaIaIaIaIaIaIaIaIaIaIa 980
Db 5730 CACTGAGCTGAAGAGCTGAAG 5671

QY 980 yEValHISgluProProArgIuAspAlaAlaProThrLysProAlaProProAlaProP
Db 5670 AAGTCCATGAGCCCCCGGGAGAGACGAGCTCCACCAAGCGCTCCCCAGCCCCAC 5611
QY 1000 roProProGlnAenLeuGlnProGlnSerAspAlaProGlnGlnProGlnSerSerProA 1020
Db 5610 CGGCACCGCAAAACCTGACACCGGAGAGACGCGCTCTGACAGAGCTGGGAGAGCCCCC 5551
QY 1020 rGgIyLysSerAArgSerProAlaProProAlaAspLysGlu----- 1033
Db 5550 GGGGCAAGACAGAGAGCCCGGACACCCCGCCGCAAGAGAGGAGAGAGAGCTGTGTTCT 5491
QY 1034 -----AlaPheAlaAlaGluAlaGlnLysLeuProGlnYAspProProCysTrpThrSerg 1052
Db 5490 TCCAGAGCTTCGACAGCGGAGCCGAGAGCTGCTGGGAGACCCCTGCTGGAATTCCG 5431
QY 1052 LysLeuProPheProAlaProProArgIuValIleLysAlaSerProHlaAlaProAsp 1072
Db 5430 GCCTGACCTTCCCGTGCCTCCCGGTGAGGTGATCAAGGCTCCCGCATGCCCCGAGAC 5371
QY 1072 roSerAlaPheSerTrpAlaProProGlnYHISProLeuProLeuGlnLysHISAspThra 1092
Db 5370 CCTACAGCTTCTCTAGGCTTCAACTGTGTCACCACTGGGCTCGGCTCATGACACTG 5311
QY 1092 LaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSera 1112
Db 5310 CCGGCGCGCTTCGCGCGCGCCACCACTCTCCAAACCGGCTCCCTCATCTCTCTG 5251
QY 1112 LaLysHISProSerValLeuGlnLysGlnIleGlnYAlaIleSergInGlnYMetSerValG 1132
Db 5250 CCAAGCACCCGACGCTCTCGAAGAGCAATAGGTGCATCTCCCAAGATGTCGTTCC 5191
QY 1132 LnduHISValProTrpTrpSergLndHISAlaLysAlaProValGlnProValThrmecGlyL 1152
Db 5190 AGCTCAAGCTCCGCTACTGACAGACTGCAAGGCGCGGTGGCGCTGATCAATGGGCG 5131
QY 1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGlnYValLysGlnGlnLnd 1172
Db 5130 TGGCCCTGCCATGAGACCCCAAAAGCTGGCACTTCACGAGAGTGAACAGAGAGAGC 5071
QY 1172 euSerProArgIyGlnIleGlnIyProProGlnSerLeuGlnYValProThraIleGlnLnd 1192
Db 5070 TGTCCCAACGGGCGCAGGCTGGGCGACCGAGAGCTGGGGGTGCCACAGCCAGAGAG 5011
QY 1192 LaSerValLeuArgIyThraIleLeuGlnYSerValProGlnYGlnSerIleThraLysGlyI 1212
Db 5010 CGTCCGTGAGAGAGGACAGCTCTGGGCTCAGTTCGCGGCGGAGAGCATCACAAAGCA 4951
QY 1212 LeProSerThraArgValProSerAspSeraIleThraTrpArgLysSerIleThrHisG 1232
Db 4950 TTCCAGACACAGGCTGCTCCGACAGCGCATCATACACCGGCTCATACCCACG 4891
QY 1232 LysThraProAlaAspValLeuYTrpLysGlyThraIleThraArgIleIleGlnYAspSerp 1252
Db 4890 GCACGCGCAGCTGAGTCTGTACAGGAGCACATCACAGATCATGGGAGAGAGAGCC 4831
QY 1252 roSerTrpLeuAspAlaArgIyAspSerLeuProLysGlnYHISValIleTrpGlnG 1272
Db 4830 CGAGTGTCTTGACCGCGCGGAGAGACAGCTGCGCAAGGCGCACGTATCTACGAG 4771
QY 1272 LysLysLysGlnYHISValLeuSerTrpGlnGlnYGlnYMetSerValThrnGlnCysSerLysG 1292
Db 4770 GCAGAGAGGCGCACGCTTGTCTTATGAGGTGGCATGTCTGTGACCCAGTCTCAAGG 4711
QY 1292 LuAspGlnYArgSerSerSergIyProProHISGlnThraAlaAlaProLysArgThra 1312
Db 4710 AGGACGGCAGAGACACTCAGAGACCCCATGAGAGCGCGCCCAAGCGCACCTATG 4651
QY 1312 apMetLeuGlnGlnYArgValGlnYArgAlaIleSerSeraIleSerIleGlnGlnYLeuMetG 1332
Db 4650 ACATGATGAGGGCGCGGTGGGAGAGCATCTCTCAGCAGCATCGAAGGTCTATGG 4591

QY 1332 LysAlaAlaIleProProGlnUArgHISerProHISISLeuLysGlnGlnYHISISleAla 1352
Db 4590 GCGGTGCATCCCGCGGAGCGACAGAGCCGCCACCACTCTAAAGACAGACACATCC 4531
QY 1352 rGgIySerIleThrnGlnGlnYIleProArgSerTrpValGlnAlaGlnGlnAspTrpLeuA 1372
Db 4530 GCGGTCTCATACACAAAGGATCCTGCTCTTACGTGAGAGGACACAGAGAGCATCTGCG 4471
QY 1372 rGArgGlnAlaLysLeuLeuLeuLysArgIuGlnIyThraProProProProProSerTrpA 1392
Db 4470 GTGAGAGGCGCAAGCTCTTAAGCGGAGAGAGAGCTCCGCCCCACCGCTTCAGCGG 4411
QY 1392 spLeuThrnGlnAlaTrpLysThrnGlnAlaLeuGlnYProLeuYLeuLysProAlaHISG 1412
Db 4410 ACCTGACCGAGGCTTCAACAGACCGAGCTCTGGGCCCCCTGAAGCTGAAGCGGCGCATG 4351
QY 1412 LngLysLeuValAlaThraValLysGlnAlaGlnYArgSerIleHISGlnIleProArgGlnG 1432
Db 4350 AGGAGCTGTGGCGCACCGGTGAAGAGGCGGCGGCTCATTCATGAGATCCCGCGAGG 4291
QY 1432 LuLeuArgHISThrnProGlnLeuProLeuAlaProArgProLeuLysGlnGlnYSerIleT 1452
Db 4290 AGCTGGGCGCACGCGCGAGCTGCTGGGCGCGCGCTCAAGAGAGGCTTCATCA 4231
QY 1452 hrGlnGlnYThraProLeuLysTrpAspThraGlnYAlaSerThrnGlnYSerLysLysHISa 1472
Db 4230 CGCAGGCGACCCCGCTCAAGTACGACACCGGCGGCTCACCATGCTGCTCCAAAGAGACG 4171
QY 1472 spValArgSerLeuIleGlnYSerProGlnYArgThraPheProProValHISProLeuAspY 1492
Db 4170 ACGTACGCTCTCTCATGCGGAGCGCGGCGGAGTTCCACCGCGTGCACCGCTGATG 4111
QY 1492 alMetAlaAspAlaArgAlaLeuGlnUArgAlaCysTrpGlnGlnYSerLeuLysSerLrp 1512
Db 4110 TGAATGCCAGCGCGCGGCACTGGAACGTGCTGTACGAGAGAGCTAAAGCGCGGC 4051
QY 1512 roGlnYThraLaseSerSergIyGlnYSerIleAlaArgGlnYAlaProValIleValProG 1532
Db 4050 CAGGAGCGCGACGACTCGGGGGGCTCATTTGCGGGCGCGCGGTGATGTGCTG 3991
QY 1532 LuLeuGlnYLysProArgGlnSerProLeuThraTrpGlnUAspHISGlnYAlaProPheAlaG 1552
Db 3990 AGCTGGGTAAAGCCGCGGAGAGCGCCCTGACCTATGAGGACACAGGCGGACCTTTGCG 3931
QY 1552 LysHISLeuProArgGlnYSerProValThrmecArgIuProThraProArgLeuGlnGlnG 1572
Db 3930 GCCACTCCACAGAGTTCGCCGTACATGCGGAGAGCCACGCGCTGAGAGAG 3871
QY 1572 LysLeuSerSerSeraLaseGlnAspArgLysLeuThraSerThraProArgGlnI 1592
Db 3870 GCAGCTTTCGTCCACAGGCAATCCAGAGCCGAAAGCTGAGTGAAGCCCTGTGAGA 3811
QY 1592 LeAlaLysSerProHISerThraIleProGlnYHISISProHISProIleSerProTrpG 1612
Db 3810 TCGCAAGTCCCGGACACAGACCGTGCAGAGACCAACCAACCCCATCTCCGCTATG 3751
QY 1612 LndHISLeuLeuArgGlnYValAspLeuYTrpArgSerHISleIleProLeuAlaP 1632
Db 3750 AGCAGCTGCTTCGGGCGGTGAGTGGGTGACCTGTATGCAACCAATCCCTGGCT 3691
QY 1632 heAspProThraSerIleProArgGlnIleProLeuAspAlaAlaAlaTrpTrpLeuP 1652
Db 3690 TCGAGCCCACTCCATACACCGCGGAGATCCCTTGAACGAGCGCTGCTTACTGCTG 3631
QY 1652 roArgHISLeuAlaProAsnProThraTrpProHISLeuYTrpProTrpLysLeuIleArgG 1672
Db 3630 CCGGACACTGGGCGGCAACCCCACTTACCGGACCTGTAACCACTCATCTGCTGCGG 3571
QY 1672 LysThraProAspHraAlaIleLeuGlnUAsnArgIlnThraIleIleAsnAspTrpIleThs 1692
Db 3570 GCTACCCCAACAGGCGGCGGTGAGAACCGGAGAGCATCATCAATGACTATCATCACT 3511
QY 1692 erGlnGlnMetHISISAsnThraAlaThraIleMetAlaGlnArgAlaAspMetLeuArgG 1712

Db	3510	CGAGACATGACCAACAACGGCCACGGCCATAGGCCACAGAGCTAATGCTGAGGG	3451
Qy	1712	lyLeuSerProArgIuSerSerLeuAlaLeuAsmTyAlaAlaGlyProArgIyIleI	1732
Db	3450	GCCTCTGCCCCGGAGGCTCTTCGTGGCACTCACTACGTGTGGGGTCCCCGAGCATCA	3391
Qy	1732	leAspLeuSerGluValProhiSleuProValleuValProProThrProGlyThrProA	1752
Db	3390	TCGACTCTCCCAAGTGGCAACCTCGCTGGCTGTGTCGCCCCCAACACAGACCCACG	3331
Qy	1752	laThrAlaMetAspArgleuAlaTyLeuProThrAlaProGluProPheSerSerArgH	1772
Db	3330	CCACGGCATGGACCGCTTGCTACTCTCCCAACGGCCCCCAGCCCTTCAGACGCCGC	3271
Qy	1772	isSerSerSerProLeuSerProGlyGlyProThrHisLeuThrIysProThrThrS	1792
Db	3270	ACAGCAGCTCCCACTCTCCCGAGAGGTCCAAACACTTACAAACCAACCAACAGT	3211
Qy	1792	erSerSerGluAlaGluIuArgAspArgAspArgIuArgAspArgAspArgIuArgIuL	1812
Db	3210	CCTGCTCCGAGGGAGGAGCGAGCCGGATCGAGAGCCGGACCCGGATCGAGCCGGANA	3151
Qy	1812	ysSerIleIeuThrSerThrThrThrValGluHisAlaProIleThrArgProGlyThrG	1832
Db	3150	AGTCATCTCTACCGTCCACCAACGAGCGGAGAGCAGCACCATCTGGAGACTGGTACAG	3091
Qy	1832	IuGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProA	1852
Db	3090	AGCAAGAGAGCGGCGAGCGGAGCGAGCGGGGGGGGGGGGAGCGAGCGAGCGGCCG	3031
Qy	1852	laSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG	1872
Db	3030	CCTCCCACTCCCATCCCAACAGCAGCCTGCCATCTCCCTCGAGCCAGAGTGGCTCC	2971
Qy	1872	InGlnArgProSerValIleuHisAsanThrGlyMetLeuGlyIleIleThrAlaValGluP	1892
Db	2970	AGCAAGAGCCCAAGTGTGCTTCCAAACAAGGATGAGGATGATATCAATCCGCTGGAGC	2911
Qy	1892	roSerIyProThrValIleuArgSerThrSerThrSerSerProValArgProAlaIat	1912
Db	2910	CCAGAGAGCCCAAGGTCTCTGAGGTCACTCTCACTCTCAACCGCTTGCCAGGTGCA	2851
Qy	1912	hrPheProProAlaThrHisCysProLeuGlyGlyIleThrLeuAspGlyValIleProThrL	1932
Db	2850	CATTCACACCTGCACCACTGCCACTGGGGGAGCACTCGATGGAGGTCTTAACCTACCC	2791
Qy	1932	eumSerGluProValIleuLeuProIyGlyAlaProArgValAlaArgProGluArgProA	1952
Db	2790	TCATGAGAGCCGTCTTGCTGGCCCAAGAGAGCCCCCGGGGTGCCCCGCGAAGCGGCC	2731
Qy	1952	rgAlaAspThrGlyHisAlaPheLeuAlaIyProProAlaArgSerGlyLeuGluProA	1972
Db	2730	GAGCAGACACCGGCGATGCTTCCCTGGCAAGCCCCCAAGCCCGGCTCGGGGTGGAGCCG	2671
Qy	1972	laSerSerProSerIySglIySerGluProArgProLeuValProProValSerGlyHisA	1992
Db	2670	CCTCCTCCCCCGAAGAGGCTTGAGGCCCCCGGCTCATGTGCTCTCTCTGCGCAGC	2611
Qy	1992	laThrIleAlaArgThrProAlaIyAsmLeuAlaProHisHisAlaSerProAspProp	2012
Db	2610	CCACCATGGCCGCAACCCCTGGAGAAAGACTGCACTCACAGCCAGCCCGACCCGC	2551
Qy	2012	roAlaProProAlaSerAlaSerAspProHisArgGlyIuIyThrGlnSerIySPropes	2032
Db	2550	CGGCGCCACCTGCTCGGCTCGGACCGCACCGGAGAAAGACTCAAGTAAACCTTTT	2491
Qy	2032	eTlIGlGluIeuGluIeuArgSerIeuGlyTyHisGlySerSerIySerProGluG	2052
Db	2490	CCATCCAGAACTGGAACCTCGATTCTTGAGGTACACGAGCGACGCTACAGCCCCGAG	2431
Qy	2052	IyValGluProValSerProValSerSerProSerLeuThrHisAspIySglIeuProL	2072

Db	2430	GGGTGGAGCCCGCTCAGCCCTGTGAGTCA	CCGAGTGTCA	CCCAAGCAAGAGGCGTCCCA	2371
Qy	2072	YHhIEuGlUGlULeUaBpLySSerHhIEuGlUGlUGlULeUaBpProLyGlnP	rog	2092	
Db	2370	AGCACCCTGGAAAGCTCGCAAGAGGCACTCGTAGGGGAGCTGCGAGCCCAAGCAG		2311	
Qy	2092	LYProValLYhIEuGlUGlUGlULaALhIEuProHsIEuAQPProLeuProGluS		2112	
Db	2310	GCCCCGTGAAGCTTGGCGGGAGGCCGCCCACTCCCAACCTCGCGCGCTGCTGAGA		2251	
Qy	2112	erGlnProSerSerSerProLeuLeuGlnInThAlaProGlyValLySGLYHhIEuGlnArgV		2132	
Db	2250	GCCACCCCTCGTCCAGCCCGCTGCTCCAAACCGCCAGGGGTCAAAAGGTCAACAGGGG		2191	
Qy	2132	aIVaLThIEuLagInHsIEuGlnValILEThrgInaBpLYrThArgHsIEuP		2152	
Db	2190	TGGTCACTCGGCGCCAGCAATAGTGAAGTATCATCAACAGACTACACCCGGCACACC		2131	
Qy	2152	roGlnGlnLeuSerAlaProLeuProAlaProLeuTYSerPheProGlyAlaSerCYsP		2172	
Db	2130	CACACAGCTTACGGCACCCCTGCGCCGCCCTTACTCTTCCCTGGGGCCAGCTGCC		2071	
Qy	2172	roValLeuApleuLArgAArgProProSerAspLeuTYrLeuProProAspHhIEuGlyA		2192	
Db	2070	CCGTCTGTGACCTCGCGCCGCCCAACAGTACTTACCTCCGCCCCGGACATAGTG		2011	
Qy	2192	LaProAlaArgLYSerProHsIEuGlnGlyLYhYsArgSerProGluProAsnLYST		2212	
Db	2010	CCCCGGCCGTGGCTCCCCCAAGAGGAAGGGGGCAAGAGTCTCCAGAGCCAAACAGA		1951	
Qy	2212	hSerValLeuGlUGlUGlUGlUaBpLYrILEuProValSerProProGluUGLYMetT		2232	
Db	1950	CGTCGTCTTGGGTGGTGTGAGAGCGGATTTGAACTGTGTCCCAACGGAGGGCATGA		1891	
Qy	2232	hGlnLProGlyHhIEuSerArgSerAlaValTYrProLeuLeuTYrArgAspGlyUGlUGInt		2252	
Db	1890	CGAAGCCAGGGCACTCCCGAAGTCTGTGTACCCGCTGCTGACCGGAGTGGGGACAGA		1831	
Qy	2252	hrgLUProSerArgMetGlySerLYSerProGlyYanThrsGlnProProAlaPheP		2272	
Db	1830	CGAGGCCAGAGAGTGGGCTCCAAGTCTCCAGGCAACACAGCCAGCCGCGACCTTCT		1771	
Qy	2272	hSerLYLeuThrgLuserAsnSerAlaMetValLYSerLYsGlnGlnILEasNL		2292	
Db	1770	TCAGCAAGCTGACCGAGAACCACTCCGCATGCTCAAGTCCAAAGAACAGATCACCA		1711	
Qy	2292	YALysLeuAnThrHsIEuArgAsnGluProGluuTYrAnLIEuSerGlnProGlyYThrg		2312	
Db	1710	AAAGCTGAACACCCCAACCCGGAATGAGCTGAAATACATATCACGACAGCTGGGACGG		1651	
Qy	2312	hILePheAsnMetProAlaILEThrgLYThrgLYLeuMetThrTYrArgSerGlnALay		2332	
Db	1650	AAATCTTCAATATGCCCCGCATACCCGGAACAGGCTTGTAGACTTATAGAACCAAGCGG		1591	
Qy	2332	aIGlGlnLysIslaSerThraMetGlyLeuGlnLysILEaLYeLYsALeUmetG		2352	
Db	1590	TCGAGGAACAATGCCAGCACCAACATATGGGCTGAGGCCATTAATGAAGACCTATGG		1531	
Qy	2352	LYLySTYrAspGlnTYrGlnGlnLYSerProProLeuSerAlaAsnAlaPheAsnProLeuA		2372	
Db	1530	GTAATATATGACCAAGTGGAAAGTCCCGCCGCTCAAGCCCAATGCTTTTAAACCTCTGA		1471	
Qy	2372	snhLAserLYAserLeuProAlaALaMetProILEThrLysALaBpGlyArgSerAspH		2392	
Db	1470	ATGCCAGTCCAGGCTCGCCCGCTCATGCCCCATTAACCGTGTGTGACGGAGGAGTACC		1411	
Qy	2392	IEThrLeuThrsErProGlyGlyGlyGlyLYsALysValSerGlyArgProSerSerA		2412	
Db	1410	AAACACTCACTCTGCCAGGTGGGGGGGAGGCCAAAGGTCTGTGGAGACCCAGCAGCC		1351	
Qy	2412	rgLYsALysErProAlaProGlyLYeuaLAserGlyAspArgProProSerValSerS		2432	
Db	1350	GAAGAACCAAGTCCCGGCGCGGCTGTGCACTGTGGAGCCGAGCACCTCTGTCTTCTCT		1291	

QY 2432 erValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluA 2452
Db 1290 CAGTGCACCTGGAGAGGAGCTGCAACCGCGGAGCGGCTGACCAACCGCTGTGGAGG 1231
QY 2452 sPaArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIlleMetArgLeuG 2472
Db 1230 ACAGGCGCTGTCTCGGAGGTTCACGCGCATTCCTCCACACCCCTGATCATGCGGCTGC 1171
QY 2472 InaIaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuA 2492
Db 1170 AGCGGGGTGATGCTTCTCCACCCCAACCGGCTCTCCGCGGAGCGGCGGCTCG 1111
QY 2492 IaGlyProHisAlaValTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluT 2512
Db 1110 CTGGCCCCCAACACGCGCTGGGAGGAGGAGCCCAAGCCACTGCTGTGCTCGACGTAAGAGA 1051
QY 2512 hIleuSerAspSerGlu 2517
Db 1050 CACTCTCCGACAGCGAG 1034

RESULT 10

US-09-819-104A-6
Sequence 6, Application US/09819104A
Publication No. US20030027137A1
GENERAL INFORMATION:
APPLICANT: Chen, J. Don
TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
FILE REFERENCE: DMG-030
CURRENT APPLICATION NUMBER: US/09/819,104A
CURRENT FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: 60/193,138
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 7386
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(7386)
US-09-819-104A-6

Alignment Scores:

Pred. No.: 0 Length: 7386
Score: 10832.50 Matches: 2149
Percent Similarity: 87.94% Conservative: 90
Best Local Similarity: 84.41% Mismatches: 197
Query Match: 81.97% Indels: 117
DB: 10 Gaps: 33

US-09-522-753-5 (1-2517) x US-09-819-104A-6 (1-7386)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 1 ATGTCAGGATCCACACAGCCTGTGGCACAACATGCGGCGCTGTGAGCCCGCTACCA 60
QY 21 ProHisSerLeuSerTyrProValGlnIlleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 61 CCCCATGCGCATCTCTCAACCGGTGCAAGATAGCCGAGTCCACACAGGACGTGGGCTGTCTT 120
QY 41 GlnTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIlleIleGln 60
Db 121 GAGTACCAACACACACCCCGGTGACTACCTCACTGTACCCGGTTCATCATTCAG 180
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 181 CCACAGAGGAGGCGGCGCTCACTGCTGTCAAGATTCCAGCTGTGGAGTGAACGGTCTCAG 240
QY 81 GlnLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
|||||
|||||

Db 241 GAGCTCCACCTGCGCGCTGAGTCCCGCAGCTTCTGCTGAGTGGCAAGCCGACATA 300
QY 101 GlnPheIlleGlySerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 301 GAATTTCACCGAGACCAAGCGCGCGCTGAGCTTACTACCGATACCTGTCTGCGCCA 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuThrLysAspArgSer 140
Db 361 TCACCTCTGCTGGCCACTGGGAGCGGAGTGGGTCTGAAGACCTTACCAAGACGCTAGC 420
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 421 CTGGCAGCAAGCTGGAGCTGTGTCACTCCAGTCCCGCAGCGCTGACCTCGAGCTA 480
QY 161 GlnLeuValProProArgLeuSerLysGluLeuIlleGlnAsn---MetAspArgVal 179
Db 481 GAGCTGGCGCATCTCTGATCTCCAGAGAGAGCTGTATCCAGAAACAGATTGGACGGGTG 540
QY 180 AspArgGluIleThrMetValGluGlnGlnIlleSerLysLeuLysLysGlnGln 199
Db 541 GACCGTGAGATCAACCATGTAGAGCAGCAGATCTCCAAAGCTGAAGAAAGCAGCAAG 600
QY 200 LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProIle 219
Db 601 TTGGAGGAGAGGCGCGCAAGCGCGCAACCGAGAAAGCTGTGTGCGCACCAACCATTA 660
QY 220 GluSerLysHisArgSerLeuValGlnIlleIleTyrAspGluAsnArgLysValGlu 239
Db 661 GAATCAAAGCAGCAAGCTGTGTCAGATCATCTACATAGAAACCGGAAGAAAGCCGAA 720
QY 240 AlaAlaHisArgLysLeuGluGluLysGlyProGlnValGluLeuProLeuTyrAsnGln 259
Db 721 GCGGCAACCGGATTCCTAGAAAGGCTGTGGGCGCCAGGTGAGAGTGTCTGTACAAACAG 780
QY 260 ProSerAspThrArgGlnTyrHisGluAsnIlleLysIlleAsnGlnAlaMetArgLys 279
Db 781 CCGCTGACACACGCGCATGTCATGAACATGAATAAACAAGCGGAGCGGAGAGAG 840
QY 280 LeuIlleuLysPheLysArgAspAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299
Db 841 CTGATCTTGACTTTAAGCGGAGAGAACACGCGCGCAAGAGTGGAAACGCGCTTGC 900
QY 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIlleGluAsnAsn 319
Db 901 CACGCGTATACACAGCTCATGAGAGCGGTGGGAGAAAGAGTACAGCCATAGAGAAACAT 960
QY 320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 961 CCGCGAAGGAGGCGCAAGGAGACCAAGGTGAGGAGTACTACGAAGAAACAGTTCCCGGAG 1020
QY 340 IleArgLysGlnArgGluLeuGlnGlnLysMetGlnSerArgValGlyGlnArgGlySer 359
Db 1021 ATCCGAAAGCGGAGGCTGTCAGAGCGCATGACAGCGAGGTGGCGCGCGGAGT 1080
QY 360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
Db 1081 GGGCTCTTCATGTGCGCTGCGCGCAGTGAAGCAAGAGTTTCTAGATCATTTAGGCTTG 1140
QY 380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeu 399
Db 1141 TCTGACAGAGAAACCTGGAGAAAGATGCGCAGCTGGCCCTGTATCC--GCCATTTTG 1198
QY 400 TyrAspAlaAspGlnGlnArgLysPheIlleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 1199 TACGATGC--GACCAGAGAGATCAAGTTTCATACATGATGACTCATGAGATGACCCC 1257
QY 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThr 439
Db 1258 ATGAAGGTCTACAAGAACCGTCAAGTTTACCAACATGTGAGCGAGCAGAGAGAGACACC 1317
QY 440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIlleAlaSerPheLeuGlu 459
Db 1318 TTCCGTGAGAGATTATGACGACCTTAAGAACTTGTGGCTGTATGCTCATATTCCTGAG 1377
|||||
|||||

460 ArgluThrValAlaGluCysValLeuTyrTyrLeuThrIleYsAsnGluAsnTyr 479
Db AGAAAGACGGTGGCTGAGTGTCTCTTACTACTGACCAAGAAAGATGAATATTC 1437
QY IysSerLeuValArgArgSerTyrArgArgGlyYsSerGlnGlnGlnGln 499
Db AAGACCTTGAGGGGAGGAGTATCGGCCCTGGCCAGACGACGACGACGACGACG 1497
QY GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Db CAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1545
QY IysAspGluIleGluYsGluYsGluIleGluYsGluIleGluYsProGluIleGlu 539
Db AAG 1605
QY AsnAspYsGluAspLeuLeuYsGluYsThrAspAspThrSerGlyIleAspAsnAsp 559
Db AACGAGAAAGAGAACTCAGCAAGAGAAAGACGACGACGACGACGACGACGACGAT 1665
QY GluYsGluIleValAlaSerYsGlyArgGlySerThrAlaAsnSerGlnGlyArgArgYs 579
Db GAGAAAGAGGCGGTGGCTTCGCAAGGCGCGCAAACTGCGACAGCGCAAGGCGCGCAAA 1725
QY GlysArgIleThrArgSerMetAlaAsnGluIleAsnSerGluIleIleThrProGln 599
Db GCGCGATACCGCGCTCCATGCGCAACGAGCCACATGAGAGACACCGCCACGACG 1785
QY GlnSerAlaGluLeuAlaSerMetGluLeuAsnGlnSerSerArgTyrThrGluGln 619
Db CAAAGTTCAGACCTGGCTTCATGAGATGAACGAGATTCGCTGGAGCTGAGAGAGAG 1845
QY MetGluThrAlaYsYsGlyLeuLeuGlnIleGlyArgAspThrSerAlaIleAlaArg 639
Db ATGAGACAGCAAG 1905
QY MetValGlySerYsThrAlaSerGlnCysYsAsnPheTyrPheAsnTyrIleYsArg 659
Db ATGGTGGGCTCCAGAGCCGTGTCCAGGTAAAGACTTCTACTTCACTCAACAAGAGAGAG 1965
QY GlnAsnLeuAspGluIleLeuGlnGlnIleYsLeuYsMetGluYsGluIleArgAsnAla 679
Db CAGAACTCGAGAGAAATCTTCAGCAGCAGCAAGCTAAAGATGAGAGAGAGAGAGAGAGAG 2025
QY ArgArgGlyLeuYsYsYsAlaProAlaAlaSerGluGluAlaIlePheProVal 699
Db CGAG 2085
QY ValGluAspGluGluMetGluIleSerGlyValSerGlyYsAsnGluGluMetValGlu 719
Db GGTGAG 2145
QY GlnAlaGluAlaLeuIleAlaSerGlyYsAsnGluValProArg--GlyGluCysSerGly 738
Db GAGGCGAAG 2205
QY ProAlaThrValAsnAsnSerSerAspThrGlnSerIlePheSerProIleThrGluAla 758
Db CCAAGCTGCTGACCAACAGCTCTGATCTGAGAGGTGCCATCCCGGCTTGAAGAGAGAG 2265
QY AlaIleYsAspThrGlyGlnAsnGlyProYsProProAlaThr-----LeuGluAlaAsp 776
Db ACGAAG 2316
QY GlyProProGluYsProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Db ACCACAGCAGCTGTCTCT-----CCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2370
QY ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
Db TCCCAAGTCCCTGATGCCAGGTGCCCAACATCCCAAGAGAGAGAGAGAGAGAGAGAGAG 2427

817 AlaProProProValValProYsGluGluYsGluGluGluThrAlaAlaProPro 836
Db GCACCCCGGCTACTGTGGACAAGAGTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2428
QY ValGluGluGluGluGluGluGluYsProProAlaAlaGluGluLeuAlaValAspThrGly 856
Db ACAGAGAGAGTCCAAAG 2538
QY IysAlaGluGluPro-----ValYsSerGlyThr 867
Db AAGCCAG 2539
QY GlnGluAlaGluGluGluYsProAla--LysGlyYsAspAlaGluAlaAlaGluAlaThr 886
Db GAGGAG 2599
QY AlaGluGlyAlaLeuIleYsAlaGluYsGluGlyYsSerGlyArgAlaThrThr--- 905
Db TGTGAGGACCACTTAAGGTGGAG-----GAGGCTGTGAGCAAGGAGAGAGAGAGAGAG 2659
QY AlaYsSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Db GGTTCAGCTCAGGTGCCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2713
QY ValAspGluAlaGluGlyGlyYsAspYsAsnArgLeuLeuSerProArgProSerLeuLeu 945
Db GTGACGAAACCGGAG 2773
QY ThrProThrGlyYsAspProArgAlaAsnAlaSerProGlnYsProLeuAspLeuYsGln 965
Db ACCCGGCTGGAGATCCCGGAG 2833
QY LeuYsGlnArgAlaAlaAlaIleProProIleGlnValThrYsValIleGluProPro 985
Db CTGAAGACAGCAG 2893
QY ArgGluAspAlaAlaProThrYsProAlaProProAlaProProProProGlnAsnLeu 1005
Db CGGAG 2950
QY GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyYsSerSerArgSer 1025
Db CAGCCAG 3010
QY ProAlaProProAlaAspYsGlu-----AlaPheAlaAla 1037
Db CCAAGTCCCTCGCGAG 3070
QY GlnAlaGlnYsLeuProGlyAspProProCysTyrThrSerGlyLeuProPheProVal 1057
Db GAGGAGCA--AAGCTACCGAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3130
QY ProProArgGluValIleYsAlaSerProIleAlaProAspProSerAlaPheSerYs 1077
Db CTTCAAG 3184
QY AlaProProGlyYsIleProLeuProLeuGlyLeuIleAspThrAlaArgProValLeuPro 1097
Db ACACCCCGGCTACCGGCTGTGCTGTGGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3238
QY ArgProProThrIleSerAsnProProProLeuIleSerSerAlaYsIleProSerVal 1117
Db CGTCCCGCC--ATCTTAACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3298
QY LeuGluArgGlnIleGlyAlaIleSerGlnYsMetSerValGlnLeuIleValProYs 1137
Db CTTGAG 3355
QY SerGlnIleAlaYsYsAlaProValGlyProValThrMetGlyLeuLeuProLeuMetAsp 1157
Db TCGAG 3415
QY ProYsYsLeuAlaProPheSerGlyValYsGlnGlnLeuLeuSerProArgGlyGln 1177

D	3472	CTTAAGAGCTG-----	3483	D	4432	AGCAGTGTCCAGGGGCTCCATCACAGTGGGCTCCAGTGTCTGTAAGTGGCC	4491
Q	1178	AlAGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly	1197	Q	1535	LyProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu	1554
D	3484	-----GGG	3486	D	4492	AAGCAGGCAAGAGCCAGTGAATTAAGAAACACAGGGGACCTTCAACAGTCACTG	4551
Q	1198	ThrAlaLeuGlySerValProGlyGlySerIleThrIleGlyLeuProSerThrArgVal	1217	Q	1555	ProArgGlySerProValIleMetArgGluProThrProArgLeuGlnGlySerLeu	1574
D	3487	ACAGGACT-GGCTCGGCACACAGTGAAGATCAACAAAGGGCT-CCAGTACCAGGGCT	3544	D	4552	CAGTGGCTCCCTGTGACACAGAGGAGGCCACGCGCCCTTCAAGAAAGCAGCTC	4611
Q	1218	ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal	1237	Q	1575	SerSerSerIleAlaSerGlnAspArgIleLeuThrSerThrProArgGluIleAlaIle	1594
D	3545	GCAGACGGGCCCC-----AGTACAGAGGCTCTATCACCAACGC-ACGCCGCGAGAGCTC	3597	D	4612	CTATCCAGCAAGGGGTCCAGGACCGGAACTGCACTACACCCCGGAGATGCCAAG	4671
Q	1238	LeuTyrIleGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg	1257	Q	1595	SerProHisSerThrValProGluHisProHisProIleSerProTyrGluHisLeu	1614
D	3598	CTCTACAAAGGGTACCATCAGAGATGCTGGTGAAGAGCCCAAGTCCGCTTGACCGG	3657	D	4672	TCCCAACAGCACTGGCCGAGACACCTCAACCACTCCCTTCAAGACACTTG	4731
Q	1258	GlyArgGluAspSerLeuProIleValIleTyrGluGlyIleValIleVal	1277	Q	1615	LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro	1634
D	3658	GCAGAGAGGACACCTCGCCCAAGGGCATGTCTATGTAGAGGCAAGAAAGCCACATC	3717	D	4732	CTCGGGGCGGTGACTGGTGTGAACCTGTACCGTGTACATCCCATTTGGACCCC	4791
Q	1278	LeuSerTyrGluGlyMetSerValThrGlnCysSerIleAspGlyArgSerSer	1297	Q	1635	ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArg	1653
D	3718	CTATCTATGAAGTGGTATGTCCGTGTACAGTGTCTAAGAGAGATGGAAGAGCAGC	3777	D	4792	ACCTCCATACCCGAGGAGATCCCTGTGAAGACAGACGCCACGCTACTGCCCCGG	4851
Q	1298	SerGlyProProHisIleGluThrAlaAlaProIleArgThrTyrAspMetMetGluIleArg	1317	Q	1654	HisLeuAlaProAspProThrTyrProHisLeuTyrProProTyrIleArgGlyTyr	1673
D	3778	TCCGGCCCAACCCATGAGACTGCGCCCTAAAGCACTATGACATGATGAGAGGGCGT	3837	D	4852	CACCTGGCCCCAGGCCCACTACCCACACTGTACCCACTTACCTCACTCGGGCTAC	4911
Q	1318	ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro	1337	Q	1674	ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln	1693
D	3838	GTAAGCAGACTGTCACTACCCAGCATAGAGGATCTATGGGCCCGCCGCTC-CT	3894	D	4912	CTGACACGGCGGCTCGAGAAACCGCAACCATCATCATATGACTATCATCCTCGCAG	4971
Q	1338	GlyArgHisSerProHisIleLeuIleGluGlnHisIleArgIleSerIleThrGln	1357	Q	1694	GlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgIleu	1713
D	3895	GAGCAGCAGACGCC---CACCTCAAGAGCAGATCATCCGAGGCTCATCAGCA	3951	D	4972	CAGATGACACCAACCTGCTCCGCGATGCCAGGCTCTGACATGCTGAGGGGCTGTG	5031
Q	1358	GlyIleProArgSerTyrValIleGlnGluAspTyrLeuArgAlaGluAlaIleValLeu	1377	Q	1714	SerProArgGluSerSerLeuAlaAsnTyrAlaAlaGlyProArgGlyIleIleAsp	1733
D	3952	GCGATCCCGAGGTCTATGTGAGGCGCAGAGAGACTTACGGCGGAGGCCAATCTC	4011	D	5032	TCAACCGAGAGTCTCGGCTGCGCCCTCAATTAATTCGCTGCGCCCAAGGACATTAAGAC	5091
Q	1378	LeuIleArgGluGlyThrProProProProSerArgAspLeuThrGluAlaTyr	1397	Q	1734	LeuSerGlnValIleProHisLeuProValIleValIleProProThrProGlyThrProAlaThr	1753
D	4012	TTBAAGCAGAAAGGACACCACTCCCAACACCACTCGGAGCTGATGAGACTTAC	4071	D	5092	CTGTCCAGATGCAACACTGCGCGTGTGTGTGCACCAACGCGACGACCCCTGCCACC	5151
Q	1398	LyThrGln-----AlaLeuGlyProLeuIleLeuIleAspProAlaHisGluGlyLeu	1414	Q	1754	AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer	1773
D	4072	AAGCCCGGCGCTGGACCTCTGGGTCCCTGAGCTGAAGCGACTCAGAGGGTGTG	4131	D	5152	GCCATGACCGCTTGTCTACTCTCCCACTGCGCCCACTCCCTTCAAGACCGCCACAGT	5211
Q	1415	ValAlaThrValIleGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg	1434	Q	1774	SerSerProLeuSerProGlyIleProThrHisIleuThrIleThrIleThrSerSer	1793
D	4132	GTAAGCACTGTGAAGAGGGCGGCGCTTATCATCAGATCCCGAAGAGGAGCTGCGC	4191	D	5212	AGCTCAACGCTGTCCCAAGAGGCCCACTACACTGTAAACCACTGCCACATCTTCA	5271
Q	1435	HisThrProGluLeuProLeuAlaProArgProLeuIleGluIleSerIleThrGlnGly	1454	Q	1794	SerGlyArgGlyArgAspArgAspArgGlyArgAspArgAspArgGlyIleIleSer	1813
D	4192	CGACACCTGAGCTACCTCGGACACAGGCTCTGAAGAGGGTTCATACACCCAGGGC	4251	D	5272	TCCGAGCGGAGACGGAACCTGAGCGGAAAGAGAC-----AGTCC	5313
Q	1455	ThrProLeuIleTyrAspThrGlyAlaSerThrThrGlySerIleIleAspValArg	1474	Q	1814	IleLeuThrSerThrThrThrValGluHisAlaProIleThrArgProGlyThrGln	1833
D	4252	ACCCCACTCAAGTACGACTGTGGGGACCTCCACTGGCACAAGAAACGAGCTGCGC	4311	D	5314	ATCTCAACGCTTACACTACAGTGAAGCATGCCATCTGGAGACTGTGTAGCGAGCAG	5373
Q	1475	SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla	1494	Q	1834	SerSerGlySerSerIleSerSerGlyGlyGlyIleIleSerSerArgProAlaSer	1853
D	4312	TTCCATATCGGAGCCCGCGCGCTTCTCCCTGCGACCCGCTGACATTAATGGCT	4371	D	5374	AGCAGCGGGGCT-----GGGGGACAGAGCGCCCGGCTCC	5409
Q	1495	AspAlaArgAlaLeuGluArgAlaCysTyrGlyGluSerLeuIleSerArgProGlyThr	1514	Q	1854	HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln	1873
D	4372	GAGCGCCGGGCACTGAGCTGCTGCTATGAAGAGCTTGAAGAGCCGCTGAGGAGCC	4431	D	5410	CACACC-----CACAGCACTGCGCATCTCCCGGAGCCAGAGCGCTTGCAGCAG	5463
Q	1515	AlaSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly	1534	Q	1874	ArgProSerValIleuHisAsnThrGlyMetIleIleThrAlaValGluProSer	1893
D	1515	AlaSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly	1534	D	5464	AGGCCAGTGTGTGCACAACAGACATGAAGGGCGTGTCTACCTCGTGAACCCGCGC	5523

1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIleThrPhe 1913
5524 AGCCCAAGCGTCTGAGGTCCACCTCCACTTCTGCTGCTCCGACGCTGCACATTC 5583
1914 ProProAlaThrHisGlySerProLeuGlyGlyThrLeuAspGlyValIleProThrLeuMet 1933
5584 CCACCTGCACACCCACGCTCCACTTGGTGACCTTGAAGGGGTCAACCTTACCTCATTAG 5643
1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAla 1953
5644 GAGCCCGTCTCTGTACCCAGAGAGACTCTCGGGTGGCCCGGCGGCGGCGGCGGCTGTG 5703
1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
5704 GAGGGGGCCATGCTTCTCCACCAAAACCCCGGCGG-----GAGCGCGCTCC 5754
1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
5755 TCACCCAGAGAGGCTCCGAGCGCCGATCCCTAGCACCCCGCCAGCTCCAGCCACAGCC 5814
1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAla 2013
5815 ATGCGCCGACCCCAAGAAAGACCTTGCACCCCACTGCTCCGAGCCGCGGCGG 5874
2014 ProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPheSerIle 2033
5875 ---CCCACTCGGCTCAGATCGACCGAGAAAGACTCAAGTAACCCCTTTCCATC 5931
2034 GlnGluLeuGluLeuArgSerLeuGlyThrHis---GlySerSerLysProGluGly 2052
5932 CAGGAATTGAAATCGGTTCTCTGGGTTCACACAGTGAAGCTGGCGCAAGCCCGATGG 5991
2053 ValGluProValSerProValSerProSerProSerLeuThrHisAspLysGlyLeuProLys 2072
5992 GTGGAGCCCATCAGCCCGGTGAGCTCCCGACGCTAACCACACAGAGGGGCTCTCCAA 6051
2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGly 2092
6052 CCTCTGGAAGAGCTTAGAAGAGCCACTTGGAGGGAGCTGGCGCAAGAGCCAGGCG 6111
2093 ProValLysLeuGlyGlyValAlaHisLeuProHisLeuArgProLeuProGluSer 2112
6112 CCATTAACCTCAGCGCGAGGCTGCCATCTCCCATCTGGCGCACCTCCCGAGAGG 6171
2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
6172 CAGCCCTCATCCAGCCCACTCTCCAGACTGCCCCAGGCAATCAAGGTCAACAGAGG 6231
2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspLysThrArgHisAsp 2152
6232 GTCAACCTGGCTCAGACATCAGAGAGTCAATACAGAGACTACACCCGGCACACCG 6291
2153 GlnGlnLeuSerAlaProLeuProAlaProLeuLysSerPheProGlyAlaSerGlyPro 2172
6292 CAGCACTCAGTGGCCCTTCCCGCCCTCTTACTCTTTCGAGAGCACTCCCT 6351
2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
6352 GTGCTGGATCTTGCGGCGCCACCAAGTACTTACTCTCCACCCCGACCACTGGCACC 6411
2193 ProAlaArgGlySerProHisSerGluGlyLysAspSerProGluProAsnLysThr 2212
6412 CCAGCCCGGGAGTCCCCCACTGAAGGGGGCAAAAGTCCCCAGAACCCAGCAAAACA 6471
2213 SerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGluGlyMetThr 2232
6472 TCGGTCTCGGCGCAGTGAAGATGCCATTGAACTGTGTGCCACAGAGGGCAATGACT 6531
2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlnThr 2252
6532 GAGCCAGACATCTCGAGCGCTGTGTACCACTGCTGTATCGAAGCGGGAGACAGGCG 6591

2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
6592 GAGCCC---AGGATGGGCTCTTAAGTCTCCAGGCAACACAGGCGGCGGCGGCTTCTTC 6648
2273 SerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysLeuGlnIleLeuLys 2292
6649 AGTAAGCTACTGAGAGCACTCCGCCATGTGTGMAATGCAAGACAGAGATCAACAAAG 6708
2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
6709 AAACCTCAACACCAACCGGAACGAGCAGCAATACATATTATGGCCAGCTGGAGAGGAA 6768
2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
6769 ATCTTCACATGCTCCCGCCATCACTGAGCAGGCGCTTATACCTGTAGAAGCCAGCGGTG 6828
2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgGlyAlaLeuMetGly 2352
6829 CAGAAACACGCGACACCAACATGGGGCTTAGAGGCCATTATTAAGAAAGCACTCATGGGT 6888
2353 LysTyrAspGlnTyrProGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
6889 AATATGATCAGTGGGAGAGCCCGCGCTCGGCGCAATGTCTTTAACTCTGAAT 6948
2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
6949 GCAGGCGCAGTGTGCTCCGCTGCTGTATGCCATTAACCACTGTGACGAGCGAGTGAAC 7008
2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyValArgProSerSer 2411
7009 CAGCACTCACTCTCGCAGGTGAGGTGAGGAAAGCCAAAGTCTTGCGAGACTAGCAGC 7068
2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
7069 CCAAAAGCCAGTGCAGACACACAGGCTAAGGCTCGGAGACCGACCCCTTCTCTCC 7128
2432 SerValHisSerGluGlyAspGlyAsnArgAlaGlyThrProLeuThrAsnArgValTyrGlu 2451
7129 TCAGTACACTCAGAGGGGAGTGCATTCGCCAAGCACCACTCAACACGTTGTGGAG 7188
2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
7189 GAGCGGCTCATCTGCAAGGTTCACGCCATTCCTTCAACCCCTTGATTAGAGGCTA 7248
2472 GlnAlaGlyValMetLysSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
7249 CAGGAGGTGTATGGCTCCCGGCCCACTGGGCTTGCGGAGGAGCGGCGGCGGCTTA 7308
2492 AlaGlyProHisHisAlaTyrAspGlyGluProLysProLeuLeuLysSerGlnTyrGlu 2511
7309 GCTGTGCCACCAACCGCTGGATGAGAGCCCAAGCACTGCTGTTCACAGTATGAG 7368
2512 ThrLeuSerAspSerGlu 2517
7369 ACACTCTCGAGACGCGAG 7386

RESULT 11
US-09-819-104A-4
; Sequence 4, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COMPRESSOR MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: UMG-030
; CURRENT APPLICATION NUMBER: US/09/819,104A
; PRIORITY FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 8544

```

; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (160)..(7545)
; US-09-819-104A-4

Alignment Scores:
Score: 0 Length: 8544
Percent Similarity: 10832.50 Matches: 2149
Best Local Similarity: 87.94% Conservative: 90
Query Match: 84.41% Mismatches: 197
DB: 81.97% Indels: 117
Gaps: 33

US-09-522-753-5 (1-2517) x US-09-819-104A-4 (1-8544)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 160 ATGTCAAGATCCACACAGCCGTGTGGCACAGACATGGCGGCTGCTGACCCGCTACCCA 219

Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 220 CCCCATGGCACTCTCTACCCGGGTGAGATAGCCCGCTCCACAGACGAGCGTGGGCTGCTT 279

Qy 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 280 GAGTACCAACACACACCCCGGACTACACCTCACCTGTACACCCGGTTCATATCCAG 339

Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 340 CCACAGAGGAGGCGGCCCTCACTGCTGTACAGATTCCAGCTGGAGAGTGAACGGTCTCAG 399

Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db 400 GAGCTTCACCTGCGCCCTGTAGTCCGACAGTTCTGCTGAGCTGGACCAAGCCGACATA 459

Qy 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 460 GAATTCACCGAGAGCAAGCCGCCGCTGGAGCTACTACCCGATACCTGTGCTGCGCCA 519

Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 520 TCACCCCTGCTGGCCACTGGGACGCGAGTGGTCTGAAGACCTTACCAAGACCGTAGC 579

Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 580 CTGGCAGGCAAGCTGAGGCTGTGTACCTCCAGTCCCGCAGCGCTGAGACCTGAGCTA 639

Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsn---MetAspArgVal 179
Db 640 GAGCTGGCGGCATCTTCACACTGTCCAGAGAGAGCTGTATCCAGAACATTTGAGACCGGCTG 699

Qy 180 AspArgGlnIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 199
Db 700 GACCGTAGATACACATGATAGACACAGATCTCCAAGCTGAAGAGAGACAGCAACAG 759

Qy 200 LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIle 219
Db 760 TTGGAGAGGAGGCGCGCCAGCGCCGGAACCCGAGAGAGCTGTGTGCCACCAACCATATA 819

Qy 220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu 239
Db 820 GAATTCAAAGCACCGAGCCGTGTCTCAATCATCTACATGAGAACCGGAAGAACCGGAA 879

Qy 240 AlaAlaHisArgIleLeuGlnGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
Db 880 GCCGCAACACCGGATCTTAGAAGGCGCTGGGCGCCAGAGTGGAGTGGCTCTGTATCAACAG 939

Qy 260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleLeuGlnAlaMetArgLysLys 279
Db 940 CCGTGTACACACCGCAGTACCATGAAACATCAAAATAAACACGAGCGATGCGGAGAGAG 999
```

```

Qy 280 LeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299
Db 1000 CTGATCTTGTACTTTAAGCGGAGGACACCGCGGCAAGCAGTGGAAACGCGCTTCTGC 1059

Qy 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319
Db 1060 CAGCGCTATGACCAAGCTCATGTAGAGCGCTGGAGAGAAAGATGAGCCGATGAGAACAAAT 1119

Qy 320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 1120 CCGGAAAGAGGCGCCAAAGAGACAGAGTGAAGATGACGAAACAGTTCGCGAG 1179

Qy 340 IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgLysSer 359
Db 1180 ATCCGACAGCAGCGGAGCTGCAGAGACGATGCAGAGCAGCGGTGGCCAGCTGGCAGT 1239

Qy 360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
Db 1240 GGGCTCTCCATGTCTGGCTGCCGAGTGAAGCATGAGGTTTCTGAGATCATTTGATGGCTTG 1299

Qy 380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db 1300 TCTGAGCAGAGAACTGGAGAAAGACAGATGCGCAGCTGGCGTGAATCC--GCCATGTG 1357

Qy 400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 1358 TAGACGCC-GACACAGCAGAGAGATCAATGTTATCAACATGATGACATCATGATGACCCC 1416

Qy 420 MetLysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThr 439
Db 1417 ATGAGAGTCTTAACAGACCCCTCAGGTTACCAACATGTGGAGCGAGCAGAGAGGACACC 1476

Qy 440 PheArgGluLysPheMetGlnHisProLysAsnPheGluLeuIleAsnPheLeuGlu 459
Db 1477 TTCGTGAGAGATTATGACGACCCCTAAGAACTTGTGGCTGTGATGCTCATTTCTTGAG 1536

Qy 460 ArgLysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db 1537 AGAAAGCGGTCCCTGAGTGTGTCTCTATTACTGACCAAGAAATGAATAATTAC 1596

Qy 480 LysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGln 499
Db 1597 AAGAGCTGTGTGAGCGGAGACTATCGCGCGTGGCAAGACGACGACGACAGCAG 1656

Qy 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlu 519
Db 1657 CAACAAACAGACAGACGACGACGAG-----ATGGCAGGAGCAGCCAGAGAGAG 1704

Qy 520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu 539
Db 1705 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1764

Qy 540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db 1765 AACGAGAGAGAGAACTCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1824

Qy 560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db 1825 GAGAAAGAGGCGCTGTGCTTCAAAAGCGCGCAAAATGTCCAAACAGCCAGGCGCGCAAA 1884

Qy 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlnGluAlaIleThrProGln 599
Db 1885 GGCCTGTATCCGCGCTCCATGGCCAAAGAGCCAAACATGAGAGAGACAGCACCCCAAG 1944

Qy 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrTrpThrGluGlu 619
Db 1945 CAAGATTCAAGAGCTGCTTCATGAGAGATGAAGAGAGATTCTGCTGAGATGAGAGAG 2004

Qy 620 MetGluThrAlaLysLysGlyLeuLeuGluLysGlyArgAsnTrpSerAlaIleAlaArg 639
Db 2005 ATGAGACAGCAAGAGAGAGGCTCTGGAACATGGAGAGAGAGAGAGAGAGAGAGAGAG 2064

Qy 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
```


DB 2065 ATGGTGGGCTCCAGACCCGTGCTCCAGTAAAGACTTCTACTTCAACTACAGAGAGAG 2124
QY 660 GlnAenLeuAspGluIleLeuGlnGlnIhIeLysLeuLysMetGluLysGluArgAnaIa 679
DB 2125 CAGAACCTGGAGAAATCTTTCAGACGACAACTTAAAGTGAAGAGAGAGAAAGCGCT 2184
QY 680 ArgArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 699
DB 2185 CCGAGAGAAAGAGAAAG 2244
QY 700 ValGluAspGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 719
DB 2245 GCTGAGAGAGAAAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 720 GlnAlaGluAlaLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 738
DB 2305 GAGGCGAAAGCTCCACAGGCTCTGGGAATGAGGTTCCAGAGTTGGGAGTGCAGTGGC 2364
QY 739 ProAlaThrValAsnaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 758
DB 2365 CCAAGCTGCTGTCAACAAAGCTCTGATTAAGAGTGTCCATCCCGCGTTCAGAAAGCC 2424
QY 759 AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAsp 776
DB 2425 ACGAAGAGAACT-----GGGCTTAAACCACTGGCACTGAGACATTGCCCCGTGCC 2475
QY 777 GlyProProGlyProProThrProProArgArgThrSerArgLysLysLysLysLysLys 796
DB 2476 ACCGACCCACCTGTTCT-----CTCCAGAAAGAACCGGACAGCCCTGCTGAGGCC 2529
QY 797 ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
DB 2530 TCCCGAGTCCCTGATCCAGATGAGCCCAACATCCCGAGAGCTTCC---CCATCAGCTGGC 2586
QY 817 AlaProProProValValProLysGluLysLysLysLysLysLysLysLysLysLysLys 836
DB 2587 GAACCCCGGCTACTGTGACAGAGATGAGCAAGAAAGCCCGGCTCTCCAGCTCCCGCAG 2646
QY 837 ValGluGluGlyGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 856
DB 2647 ACAGAAAGATGCCAAG 2697
QY 857 LysAlaGluLysPro-----ValLysSerGlyCysThr 867
DB 2698 AAGCCAGAGAGAGCCGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2757
QY 868 GluGluAlaGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 886
DB 2758 GAGGAGACCGAG 2817
QY 887 AlaGluGlyValAlaLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 905
DB 2818 TCTGAGGACCACTTAAAGTGGAG-----GAGGCTGTGAGCAAGGACGCTGACCAAG 2871
QY 906 AlaLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 925
DB 2872 GGTTCAGGCTCAGAGTGCACCCAGGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2931
QY 926 ValAlaAspGluAlaGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 945
DB 2932 GTGGAG 2991
QY 946 ThrProThrGlyAspProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 965
DB 2992 ACCCGGCTGGAGATCCCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3051
QY 966 LeuLysGlnArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 985
DB 3052 CTGAAGCAG 3108
QY 986 ArgGluAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1005

DB 3109 CCGAGAGACAGTACCCCAAGGACGTTCCCTGTGCTTCACCCAGAGAGAGAGAGAGAG 3168
QY 1006 GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer 1025
DB 3169 CAGCCAG 3228
QY 1026 ProAlaProProAlaAspLysGlu-----AlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1037
DB 3229 CCAAGTCCCTCTCCGAG 3288
QY 1038 GlnAlaGlnLysLeuProGlyLysProProCysTrpThrSerGlyLysLeuProPheProAla 1057
DB 3289 GAGGAGCA-AGCTACCAAGTGAAGCCCAAGCTGTATTCGGCT---GCTTTCAT--- 3342
QY 1058 ProProArgGluValIleLysAlaSerProAlaAlaProAspProSerAlaPheSerLys 1077
DB 3343 CCTCCACGGAGAGTGTACAAAGCTTCCACAGCGCT---GACCT---CTTTCTCTCTAC 3396
QY 1078 AlaProProGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1097
DB 3397 ACACCCCGGAGTACCCGCTGCTCTGGGCTTCAGAGATGAGTGGCCGCGCTGCGCA 3456
QY 1098 ArgProProThrLysSerAsnProProProLysLysSerSerAlaLysLysProSerVal 1117
DB 3457 GGTCCCCC---ATCTTAACCCCAACCCCTCATCTCTCTGCAAGATCCCGGCGTA 3513
QY 1118 LeuGluArgGlnIleGlyAlaIleSerGlnLysMetSerValGlnLeuLysValProLys 1137
DB 3514 CTTGAGAGAGCAGCTGGTGCATCTCCAGAGGAGTGTAGTCAAGCTTGCTGTGCTTCC 3573
QY 1138 SerGluLysAlaLysAlaProValGlyProValThrMetGlyLysLeuProLysProMetLys 1157
DB 3574 TCAGAGCATGCCAAG---CCATGGGCTCTTCACCATGAGCTGCTTCGCGTGAAC 3630
QY 1158 ProLysLysLeuAlaProPheSerGlyValLysGlnGluLysLeuSerProArgLysGln 1177
DB 3631 CTTAAGAGCTG----- 3642
QY 1178 AlaGlyProProGluSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1197
DB 3643 -----GGG 3645
QY 1198 ThrAlaLeuGlySerValProGlyLysLysLysLysLysLysLysLysLysLysLysLys 1217
DB 3646 ACAGCACT-GGCTCCGCCACAGTGAAGATCACCAGAGGCT-CCAGTACCCCGGCT 3703
QY 1218 ProSerAspSerAlaIleThrTyraArgLysSerLysLysLysLysLysLysLysLysLys 1237
DB 3704 GCAAGAGGCCCC-----AGCTACAGAGCTTATTCACCAAGC-ACGCCCGCAGAGCTC 3756
QY 1238 LeuTyLysGlyThrLysThrArgLysLysLysLysLysLysLysLysLysLysLysLysLys 1257
DB 3757 CTCTACAAAGGATACATCAAGAGATGTCGGGTGAGAGACAGCCAAAGTGGCTTGACCGG 3816
QY 1258 GlyArgGluAspSerLeuProLysGlyLysValIleTyraGluLysLysLysLysLysLys 1277
DB 3817 GCACGAGAGAGACCTGTCACAAAGGCTCATGATCAAGAGAGAGAGAGAGAGAGAGAG 3876
QY 1278 LeuSerTrpGluGlyGlyLysMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
DB 3877 CTATCTATGAAGGATGATGTCGTGTACAGTGTCTTAAGAGAGATGGAAGAGAGAGAG 3936
QY 1298 SerGlyProProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1317
DB 3937 TCGGGCCCAACCATGAGACTGCGCCCTTAACGCACTTAAGACATGATGAGAGGCGCGT 3996
QY 1318 ValGlyArgAlaIleSerSerAlaSerLysLysLysLysLysLysLysLysLysLysLysLys 1337
DB 3997 GTAGGAGAGACTGTCACTCAAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4053
QY 1338 GlnArgLysSerProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1357
DB 4054 GAGCAGCAGAGCCCC---CACTTCAAGAGAGAGAGATCAATCGAGGCTTCATCAGCGAA 4110

Qy 1358 G1Y1LeProArgSerTYrValGluAlaGlnGluAspTYrLeuAArgGluAlaLeuLeu 1377
Db 4111 GGATCCCGAGAGTCTCTATGTGGAGGCGAGAGGACTTACCGCGGAGGCCAAGTCC 4170
Qy 1378 LeuLeuArgGluGlyThrProProProProProSerArgAArgLeuThrGluAlaTYr 1397
Db 4171 TTGAAGGAGAAAGGAGACCACTCCCGACACCACTCCGGAGACTGAGACCACTAC 4230
Qy 1398 LyeThrGln-----AlaLeuGlyProleuYsleuYsProAlaHisGluGlyLeu 1414
Db 4231 AAGCCCCGCGCCCTGAGACCTCTGAGTCCCTGAGAGCGAAGCCGACTCAGAGAGGTGTG 4290
Qy 1415 ValAlaThrValLysGluAlaGlyArgSerLleHisGluLleProArgGluGluLeuArg 1434
Db 4291 GTAGCACTGTGAAGAGAGCGCGCGCTCTATCCATGAGATCCCGAGAGAGACTGCGC 4350
Qy 1435 HisThrProGluLeuProLeuAlaProArgProleuYsGluGlySerLleThrGlnGly 1454
Db 4351 CGCACACCTGAGCTACCCCTGGCACCAAGGCGCTGGAAGAGGGTTCCATCACCGAGGC 4410
Qy 1455 ThrProleuYsTYrAAspThrGlyAlaSerThrThrGlySerLysLysHisAAspValArg 1474
Db 4411 ACCCCACTCAAGTACGACTGTGGGGCACCTCCACTGGCACCAAGAAACAGAGTGGCG 4470
Qy 1475 SerLeuLleGlySerProGlyArgThrPheProProValHisProLeuAAspValMetAla 1494
Db 4471 TCCATCATCGGAGAGCCCGCGCGGCTTCCCTCCCTCCGACCGCGCTGACATATAGCT 4530
Qy 1495 AspAlaArgAlaLeuGluArgAlaCysTYrGluGluSerLeuYsSerArgProGlyThr 1514
Db 4531 GAGCGCGCGGACACTGAGCGGTGCTGCTATGAAGAGTCTGAAGAGCGGCTCAGGAGCC 4590
Qy 1515 AlaSerSerSerGlyGlySerLleAlaArgGlyAlaProValLleValProGluLeuGly 1534
Db 4591 AGCAGTGGTCAAGGGGCTCATCACAGTGGGGCTCAGTCGTCGCTGAGACTGGGC 4650
Qy 1535 LysProArgGlnSerProLeuThrTYrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db 4651 AAGCCAGCGAAAGCCCACTGACTTACGAAAGACACCGGGGACCTTACACAGTCACTG 4710
Qy 1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeu 1574
Db 4711 CCAGTGGTCTCCCTGTGACCAAGAGGAGCCACGCGCTTCAAGAAAGGAGGCTC 4770
Qy 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluLleAlaLys 1594
Db 4771 CTATCCGAGAAAGGCTCCAGAACCGGAACTGACATCTACACCCCGGAGATCGCCAG 4830
Qy 1595 SerProHisSerThrValProGluHisHisProHisProLleSerProTYrGluHisLeu 1614
Db 4831 TCCCCACACAGACTGTGCCGAGACCAACCTCACCTCCCTTATGACACTTG 4890
Qy 1615 LeuArgGlyValSerGlyValAspLeuTYrArgSerHisLleProLeuAlaPheAspPro 1634
Db 4891 CTCGGGGGCGTGACTGGTGTGACCTGTACCGTGTACATCCATTTGGCTTTGACCCC 4950
Qy 1635 ThrSerLleProArgGlyLleProLeuAsp---AlaAlaAlaAlaTYrTYrLeuProArg 1653
Db 4951 ACCTTCATACCCGAGGATCCCTCTGAAAGACGACCCGAGCTTCTACTGCCCCGG 5010
Qy 1654 HisLeuAlaProAspProThrTYrProHisLeuTYrProProTYrLeuLleArgGlyTYr 1673
Db 5011 CACTGGCCCCCGACCCCACTTACCACTGTACCTTACCTTACCTTACCTGAGGCTAC 5070
Qy 1674 ProAspThrAlaAlaLeuGluAspArgGlnThrLleLleAspAspTYrLleThrSerGln 1693
Db 5071 CTTGACACGGCGCGCTGAGAAACCGCAATCATCATATGACTCATCATCTCCAG 5130
Qy 1694 GluMetHisHisAspThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 5131 CAGATGACACCAACGCTGCTCCGCTGATGACCGACGCTGACATGCTGAGGGGTCTG 5190

Qy 1714 SerProArgGluSerSerLeuAlaLeuAspTYrAlaAlaGlyProArgGlyLleLleAsp 1733
Db 5191 TCACCGGAGAGTCTCTGCTGGCCCTCATATTATTCGCTGGCCCAAGAGCATTATGAC 5250
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db 5251 CTGTCCCAAGTGCACACCTGCGCTGTGTGCACCAACGCGAGGACCCCTGCGCAC 5310
Qy 1754 AlaMetAspArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 5311 GCATTCAGCGGCTTGCTTACCTTCCACTGCGGCCCAACCTTTCAGACGCGCCACAGT 5370
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793
Db 5371 AGCTCACGGCTGTCCAGAGAGGCCCACTTCACTTAAACCACTGCACACTTCA 5430
Qy 1794 SerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluYsSer 1813
Db 5431 TCGGAGCGGGAACGGAACTGAGCGGGAACGAGAC-----AAGTCC 5472
Qy 1814 LlleuThrSerThrThrThrValGluHisAlaProLleTPArgProGlyThrGluGln 1833
Db 5473 ATCTTCAGTCTACCTACAGTGAAGCATGACCTCATGTGAAGCTGTGATCGAGACG 5532
Qy 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerSerArgProAlaSer 1853
Db 5533 AGCAGCGGGCT-----GGGGGAGACACCGCCGCTCC 5568
Qy 1854 HisSerHisAlaHisGlnHisSerProLleSerProArgThrGluAspAlaLeuGlnGln 1873
Db 5569 CACAC-----CACAGCACTGCGCCATCTCCCGCGAGCCAGAGCGCTTGACAG 5622
Qy 1874 ArgProSerValLeuHisAspThrGlyMetLysGlyLleLleThrAlaValGluProSer 1893
Db 5623 AGCGCCAGTGTGTGCAACACGAGCATGAAGGCGTGTCATCTCGTGAACCGGCG 5662
Qy 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
Db 5683 AGCGCCAGGCTCTGAGTCACTTCACTTCCCTTCCGCTCCGCGAGCTGCACATTC 5742
Qy 1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTYrProThrLeuMet 1933
Db 5743 CCACCTGCCACCACTGCCCACTTGTGGCACTTGAAGGGGTCTACCTTACCCTCATG 5802
Qy 1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAla 1953
Db 5803 GAGCCCGTCTGTACCAAGAGACCTCTGAGTGCCTCGCGCGAGCGGCGCGCTGTG 5862
Qy 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Db 5863 GACGATGGCATGCTTCTTCTCACCAACCCCGGCGG-----GAGCCGCTCC 5913
Qy 1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Db 5914 TCACCCGAGAAAGTCTCGAGGCCCGATCCCTAGACCCCGGAGCTCAGACACAGACC 5973
Qy 1994 LlleAlaArgThrProAlaLysAspLeuAlaProHisHisAlaSerProAspProProAla 2013
Db 5974 ATGCGCCGACCCCGCAAGAGCTTGCACCCCAACATCCAGTCCGAGCCGCGGGG 6033
Qy 2014 ProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerLle 2033
Db 6034 ---CCACCTCGGCTCATGATCTGCACCGAGAAAGACTCAAGATAACCTTTTCATC 6090
Qy 2034 GlnGluLeuGluLeuArgSerLeuGlyTYrHis---GlySerSerTYrSerProGluGly 2052
Db 6091 CAGGAATTGGAATCTCGTCTCTGAGTACCAAGTGAAGTGTGACACCCCGATGGG 6150
Qy 2053 ValGluProValSerProValSerSerProSerLeuThrHisAspArgLysGlyLeuProLys 2072
Db 6151 GTGAGGCGCATTCAGCGCGGTGAGCTCCCGAGCTTACCAACGAGAGGGGCTCTCCAA 6210
Qy 2073 HisLeuGluGluLeuAspArgLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092

```
Db 6211 CCTCTGGAAAGAGTGAAGAGCCACTTGGAAAGGAGCTGGGACACAGCAGCCAGGCG 6270
Qy 2093 ProvaliylsleuGlyylgluaiaahisieuProhi:sleuargProleuprogIuser 2112
Db 6271 CCATGAAACCTACAGCCGAGGAGGTGGCCATCTCCCACTTGCGGCACTGGCCGAGAGC 6330
Qy 2113 GlnProSerSerSerProleuLeuGlnThrAlaProGlyValIysGlyIysGlnArgVal 2132
Db 6331 CAGCCCTCATCCAGCCACTCTCCAGACTGGCCCGAGCATCAAGGTCCACGAGAGGTG 6390
Qy 2133 ValThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspTyrThrArgHisIlePro 2152
Db 6391 GTACCCCTGGCTCAGACATCAGCGAGGTCTATTCAGAGATCAACCCGCGACCCGCG 6450
Qy 2153 GlnGlnLeuSerAlaProleuproAlaProleuTyrSerPheProGlyAlaSerCysPro 2172
Db 6451 CAGCACTAGTGGGCCCCCTTCCCGCCCTCTTACTCTTCTTCCGAGAGCCAGTCCCT 6510
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuTyrIleuProProAspHisGlyAla 2192
Db 6511 GTGCTGTGATCTTGGCGCCACCCAGTGAACCTTACCTCCACCCCGCAGCATGGCACC 6570
Qy 2193 ProAlaArgGlySerProHisSerSerGlyGlyIysArgSerProGluProAsnIlyThr 2212
Db 6571 CCAGCCCGGGGATCCCCCACTGAAGGGGCGMAAGTCCCAAGACCCAGCAAAACA 6630
Qy 2213 SerValLeuGlyIylgluaasgIyllegIuProValSerProProGluGlyMetThr 2232
Db 6631 TCGGTCTTGGGCGAGCTGAGATGCGATTCGACCTGTGTCCCAACAGAGGCACTGACT 6690
Qy 2233 GluProGlyHisSerArgSerAlaValTyrProleuLeuTyrArgAspGlyGlnThr 2252
Db 6691 GAGCCAGGACATGCTCGAGCGCTGTGTACCACTGCTGTATGAGACGGGGAACAGGGC 6750
Qy 2253 GluProSerArgMetGlySerIysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Db 6751 GAGCCC--AGGATGGGCTCTAAGTCTCCAGGCAACAGCCAGCGCGCCGCTTCTTC 6807
Qy 2273 SerIysLeuThrGluSerAsnSerAlaMetValIysSerIlyIysGlnIleAsnIly 2292
Db 6808 AGTAACTGACTGAGAGCAATCTCGCCATGCTAAGTCCAGAAAGAGAGATCAACAG 6867
Qy 2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
Db 6868 AAACCTCAACCCCAACCGGAAAGAGCCAGATTAATTTGGCCAGCTGGAGCGAA 6927
Qy 2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 6928 ATCTTCAAATGCTCCCGCATCTGAGAGCAGGCGCTTATGACTGTGAAAGCCAGGCGGTG 6987
Qy 2333 GlnGlnHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgIysAlaLeuMetGly 2352
Db 6988 CAAAGAACGCGCCAGCAACATGGGGCTTGAAGGCCATTTTGAAGAGCACTCAATGGGT 7047
Qy 2353 LysTyrAspGlnIleProGluIuserProProLeuSerAlaAsnAlaPheAsnProleuAsn 2372
Db 7048 AATATGATGATGAGGGAAGAGCCCGCGCTCGGCGCAATGCTTTTAACCTCTGAAT 7107
Qy 2373 AlaSerAlaSerIleuPro--AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 7108 GCGAGGCGCATGTGCGCTGCTGTATGCCATTAACCACTGCTGAGCGAGCGAAGTGAC 7167
Qy 2392 HisThrLeuThrSerProGlyGlyIylgluaIlyValIysValSerGlyArgProSerSer 2411
Db 7168 CAGGCACTCACTTGGCGAGGTGGAGAGGCAAGGTCTTGGCAACCTTAGAGGC 7227
Qy 2412 ArgIysAlaIysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 7228 CGAAAAGCCAAAGTGGCAGCAGCAGGCTTATGCGTCCGAGAGCCGAGCCCTTCTGTCTCC 7287
Qy 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProleuThrAsnArgValITProGlu 2451
```

```
Db 7288 TCAGTAACCTCAGAGGGGAGCTGCAATGCGCGAACAACCACTACCAACCCGTGTGGAG 7347
Qy 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProleuIleMetArgLeu 2471
Db 7348 GACCGGCCCTCATCTGAGAGGGTCCAGCGCATTCCTCCCAACCCCTTGTATTAAGAGCTA 7407
Qy 2472 GlnAlaGlyValMetAlaSerProProProGlyIleuProAlaGlySerGlyProleu 2491
Db 7408 CAGGAGGGGTATGCTCCCGCCCGCCAGCTGGCTTGGCGAGGAGCGGCGCCCTTA 7467
Qy 2492 AlaGlyProHisIleAlaThrAspGluGluProIysProleuLeuCysSerGlnTyrGlu 2511
Db 7468 GCTGTGCCCAACACGCTGGGATGAGAGCCCAAGCCACTGCTGTTCACAGTATGAG 7527
Qy 2512 ThrLeuSerAspSerGlu 2517
Db 7528 ACACCTCGGACAGCGAG 7545

RESULT 12
US-10-764-420-1526
; Sequence 1526, Application US/10764420
; Publication No. US20050084872A1
; GENERAL INFORMATION:
; APPLICANT: Lum, Pek Yee
; APPLICANT: Tan, Yeyun
; APPLICANT: Dai, Hongyue
; TITLE OF INVENTION: Methods For Determining Whether An Agent
; TITLE OF INVENTION: Possesses A Defined Biological Activity
; FILE REFERENCE: ROSA122057
; CURRENT APPLICATION NUMBER: US/10/764,420
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/442,797
; PRIOR FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US 60/474,413
; PRIOR FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 3683
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1526
; LENGTH: 8544
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-764-420-1526

Alignment Scores:
Pred. No.: 0 Length: 8544
Score: 10832.50 Matches: 2149
Percent Similarity: 87.94% Conservative: 90
Beet Local Similarity: 84.41% Mismatches: 197
Query Match: 81.97% Indels: 117
DB: 22 Gaps: 33

US-09-522-753-5 (1-2517) x US-10-764-420-1526 (1-8544)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnITrPArgAlaThrGluProArgTyrPro 20
Db 160 ARGTCAGATCCACACAGAGCTGTGGCAACACATGGCGGGCTGTAGCCCGCTACCCA 219
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 220 CCCCATGGCATCTCCATCCCGGTGAGATAGCCGGTCCCAACGGAACGGGGGCGCTT 279
Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 280 GAGTAAACAACACACACCCCGTGTACTACACTCAACACTGTCAACCCGGTTCATCATCCAG 339
Qy 61 ProGlnArgArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 340 CCAACAAGAGAGGGGCGCTCACTGCTGTCAAGATTCCAGCTGGGAGTGAACGGTCTCAG 399
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrIleuProGluIylgluaIlyIysSerGluMet 100
Db 400 GAGCTCACTGCGCCCTTAAGTCCCGCAGCTTCTGTGAGTCTGGGCAAGCCGACATA 459
```

OY		101	G u p h e e g i u s e r y a r p r o a r g l e u l e u P r o A s P r o l e u L a r P r o	120
Db		460	G A T T C A C G A G A G C C C C C G G A G T A C T A C C G A T A C C T G T G G C C A	519
OY		121	S e r P r o l e u L a r h r c y g i n P r o a a g y S e r G u A s P r e u L e u Th r y A s P r S e r	140
Db		520	T A C C C T G T G C C A C T G G C A C C G A T G G T C T G A A G A C T T A C A A G A C G T A C	579
OY		141	L e u Th r G y L y L e u L u P r o V a I S e r P r o S e r P r o H a t H r A s P r o J u L e u	160
Db		580	C T G C A G C A G C T G A G C T G T G C A C T C C A G T C C C C G A C G T G A C C T G G T A	639
OY		161	G U L e u V a I P r o P r A r g L e u S e r y S e r G U L e u L e g i A s N -- M e t A s P r y A	179
Db		640	G A G C T G G C C C A C T C T G A C T G T C A A G G A G G A G C T G A A C A G A T T G A C C G C G T	699
OY		180	A A P r A r G L u I e Th r M e T A G U G i n G i n I e S e r L y L e u L y L y S e r G i n G i n	199
Db		700	G A C C G T G A T A C C A T G T A A G A C A G A C T C C A G C T A A G A A G A A G A C A A G	759
OY		200	L e u G i U G i U A I A L A L y S P r o J u P r o J u L y P r o V a I S e r P r o J u L e u	219
Db		760	T T G A G A G A G A G C C G C C A G C G C C C A C C G A A C C T G T G C G A C A C C A T A	819
OY		220	G U S e r L y H A r S e r L e u V a I G i L e I e r y A S P r G U A S A r G i L y S A I a G U	239
Db		820	G A A T C A A G A C C G A A G C C T G T C C A G A T C A T C T G A T G A A A C C G G A A C C G A	879
OY		240	A A A A H A r G i L e u G i U L y L e u G i P r o G i N V a I G U L e u P r o L e u Th r S e r G i N	259
Db		880	G C C G A C A C C G G A T C T G A A A G C C T G G G G C C C A G T G A C T G T G T A C A A C A G	939
OY		260	P r o S e r A S P r H r A r G i N T r h S G U A S i N I e L y I e A r G i N A I A M e r A r G y S y S	279
Db		940	C C G T G A C A C A C G C A G T A C A T A A A A C A A A T A A C A G C A T G G G A A G	999
OY		280	L e u I e L e u Th r P h e L y A r G r A r S i S A I A r G y S G i N T r P L y S G i N L y S h e C y S	299
Db		1000	C T G A C T T G A C T T T A A G G A G A A C C A C G C C C A G A C A G T G A A A C A G C T T G C	1059
OY		300	G i N a r G T r A r G i N L e u M e C i U A I A L e u G i U L y L y S V a I G U A r G i E G i U A S i N	319
Db		1060	C A G C C T T G A C A C A C T C T G A G G C G T G G G A A A G T A G A C G A T A G A A A C A T	1119
OY		320	P r O r A r G r A r A L y S G i U S e r y S V A A r G y U r T r Y r G i U L y S G i P h e r P r o J u	339
Db		1120	C C G C A A G A G G C C A G A G A G C A A G T G A G G A G T A C A A A A C A G T T C C G G A G	1179
OY		340	I L e r G y S G i N A r G i U L e u G i N U r G M e G i N S e r A r G V a I G i N A r G y S e r	359
Db		1180	A T C C C A G A C A G C G G A G C T G C A G A G G C A T G C A G A G G T G G C A G T G G C A G T	1239
OY		360	G i L y L e u S e r M e S e r A I A A r S e r G i U h i S G i U V a I S e r G i U i L e A S P i L e u	379
Db		1240	G G G C T C C A G T C G G T G C C G C G A G T G A G C A T G G A T C A T G A T G G C T T G	1299
OY		380	S e r G i U G i N U A S i N L e u G i U L y S G i N M e r A r G i N L e u A I A V a I I P r o B r o k e r L e u	399
Db		1300	T C T G A G C A G A C C T G A A G A C A T G C C C A G C T G G C C G T A T C-- G C A T G T G	1359
OY		400	Tyr A s P r A I A S P r G i N G i N A r G i L e y S h e I e A S i N M e A S i N G i L y L e u M e r A I A S P r o	419
Db		1358	T A C A G C G-C A C A G C A G G A G T C A A G T T C A T C A A C A T G A T G A C T C A T G A T G A C C C	1416
OY		420	M e L y S V a I T r L y A S P r A r G i N V a I M e A i N M e	

Db	1537	AGAAAGACGGTCGCTGAGGTGTCTCTTAATTACTACCTGACCAAGAAATGAAATTAAC	1596
Qy	480	LySserLeuValArgArgSerTyrArgArgGlyLySserGlnGlnGlnGln	499
Db	1597	AAGACCTTGTAAGCGGAGACTATGGCGCCCTGGCAAGACGACGACGACGACG	1655
Qy	500	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	519
Db	1657	CAACAACAGACAGCACACAGCAG-----ATGGACCGAGCAGCCAGGAGG	1704
Qy	520	LySArgGlyLySArgGlyLySArgGlyLySArgGlyLySArgGlyLySArgGly	539
Db	1705	AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1766
Qy	540	AenAapLySArgGlyLySArgGlyLySArgGlyLySArgGlyLySArgGlyLyS	559
Db	1785	AACGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1822
Qy	560	GlyLySArgGlyLySArgGlyLySArgGlyLySArgGlyLySArgGlyLyS	579
Db	1825	GAGAAAGAGGCGCGTGGCTCCAAAGGCGGCAAACTGGCAACAGCAGGCGCG	1888
Qy	580	GlyArgGlyLeuThrArgSerMetAlaAsnGlyAlaAsnSerGlyAlaLeuThr	599
Db	1885	GGCCCTTATCAGCGCTGCATGCGCCCAACGAGCCCAACTGTAGGAGACGCCA	1944
Qy	600	GlnSerAlaGlyLeuLeuAlaSerMetGlyLeuLeuGlnGlnGlnGlnGln	619
Db	1945	CAAACTTCAAGACTGGCTTCCATGGAGATGAACGAGATCTTCGCTGACATGA	2004
Qy	620	MetGlnThrAlaLeuLySArgGlyLeuLeuGlnGlnGlnGlnGlnGlnGln	639
Db	2005	ATGGAGACAGCAAGAAAGGCTCTCGAATCGGAGAACTGGTACGCAATGCCGC	2066
Qy	640	MetValGlySerLeuThrValSerGlnCysAlaAsnPheTyrPheAsnTyrLyS	659
Db	2065	ATGTGGGCTCCAAAGCCGTGCTCCAGTGAAGAACTTCACTTCACTCAAGAA	2124
Qy	660	GlnAenLeuAapGlyLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	679
Db	2125	CAGAACTCGACGAAATCTTTCACAGACACAAGCTAAAGATGGAGAAAGAAAG	2188
Qy	680	ArgArgLySLeuLySLeuAlaProAlaAlaAlaSerGlnGlyAlaAlaPhePro	699
Db	2185	CGAGGAAAGAAAGAAAGAACCCCAAGCTCGCGCGAGCGAGGAGACAGCTTCC	2244
Qy	700	ValGlnAapGlnGlnGlnMetGlyAlaSerGlyValSerGlyLeuGlnGlnGln	719
Db	2245	GCTGAGGACGAAAGATGAGGATGACATGAGCCCAAGTGCATGTAGAAAGACT	2304
Qy	720	GlyAlaGlyAlaLeuAlaLeuAlaSerGlyAlaAsnGlyAlaProArg---Gly	738
Db	2305	GAGGAGAGGAGCTCACAGGCTCTGGAGATAGGTTCACAAGTTGGGAGATGAG	2364
Qy	739	ProAlaThrValAlaAsnSerSerAspThrGlySerLeuProSerProAlaThr	758
Db	2365	CCAGTGTCTGCAACAACAGCTCTGATTACTAGAGTGTCCATCCCGCGTTCAAG	2424
Qy	759	AlaLySArgThrGlyGlnAenGlyProLySArgProAlaThr-----Leu	776
Db	2425	ACGAAAGACACT-----GGGCGTAAACCACTGGACATGAAGCATTCCGCGT	2475
Qy	777	GlyProProProGlyProProThrProProArgArgThrSerArgAlaProL	796
Db	2476	ACCAAGCCACTGTATCTT-----CTTCAAGAAACCGGAGAGCCCTCGTAGCC	2529
Qy	797	ThrProAlaSerGlyAlaAlaThrGlyAlaProThrProProProAlaProSer	816
Db	2530	TCCCCAGTCCCTGATGCCAGTGGCCACATCCCAAGACCTTCC---CATATCCT	2586
Qy	817	AlaProProProValProLyGlnGlnGlyLySArgGlnGlnGlnGlnGlnGln	836

[illegible]

QY	1178	AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly	1197
Db	3643	-----	-----Ggg 3645
QY	1198	ThrAlaLeuGlySerValProGlyGlySerIleThrIlyGlyIleProSerThrArgVal	1217
Db	3646	ACAGCACT -GGCTCGGCACACAGTGAAGATCATCCAAAGGGCTT -CCAGTACC GGACT	3703
QY	1218	ProSerApSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal	1237
Db	3704	GCAGACGGCCCC-----AGCTAACAGAGGCTCTTACCCACCG -ACGCCCGAGACGTC	3756
QY	1238	LeuTyrIlysgIyThrIleThrArgIleIleGlyIleuAspSerProSerArgLeuAspArg	1257
Db	3757	CTTACAAAGGATACATCAGACAGATGCTCGTATGAGACACGCCCAAGTCGCTTGACCGG	3816
QY	1258	GlyArgIleuAspSerLeuProGlyHisValIleTyrGlnIlyIlyblyGlyHisVal	1277
Db	3817	GCAGAGAGGACACCTCGCCCAAGGGCCATCTCATATGAGGCGCAAGAAAGCCACGTC	3876
QY	1278	LeuSerTyrGlnIyGlyMetSerValThrGlnIySerIlysgIyIleuAspArgIlyArgSerSer	1297
Db	3877	CTATCTTATGAAGGGATGATGCTCGTGCACAGTCTCTTAAAGAGATGGAAGACAGC	3936
QY	1298	SerGlyProProHisGlyuThrAlaAlaProIlybArgThrTyrAspMetMetGlyIlyArg	1317
Db	3937	TGGGGCCACCCATGAGACTGCGCCCTTAAAGCACTATGATCATGTAGAGGGCGGT	3996
QY	1318	ValGlyArgAlaIleSerSerHisSerIleGlnIyIleuMetGlyArgAlaIleProPro	1337
Db	3997	GTAGCGAGAGCTGACCTCGACCTGACAGCATAAAGGACCTCATGGGCGCGGCATC - -CCT	4053
QY	1338	GlyuArgHisSerProHisHisIleuIlyGlnIlyHisIleArgGlySerIleThrGln	1357
Db	4054	GAGCAGCACACCCCC---CACCTTAAGAGGAGCATCATCCGAGGCTCATCACCGAA	4110
QY	1358	GlyIleProArgSerTyrValGluAlaGlnIleuAspTyrIleuArgGlyAlaIlyLeu	1377
Db	4111	GGCATCCCGAAGTCTCTATGTGAGCGGAGAGGACTCATTCAGCGCGGAAGCCAAAGCTC	4170
QY	1378	LeuIlyAspGlnIyThrProProProProProSerArgAspLeuThrGluAlaTyr	1397
Db	4171	TTGAAGCGAAGAGGACACACCTCCCCACACACACTCGGAGACTGATCAGAACCTAC	4230
QY	1398	LysThrGln-----AlaLeuGlyProLeuIlyLeuIlyProAlaHisGlyIleu	1414
Db	4231	AAAGCCCGGCCCCCGAGACCTCTGGGTCTCCCTGAAGCTGAAGCGGACTCACAGAGGGTGTG	4290
QY	1415	ValAlaThrValIyGlyAlaGlyIlyArgSerIleHisGluIleProArgGlnIleuArg	1434
Db	4291	GTAGCACTGGAAGAGAGCGGGCGGCTTATCATAGATCCGAGAGAGAACTCGC	4350
QY	1435	HisThrProGlnIleuProLeuAlaProArgProLeuIlyGlnIySerIleThrGlnIy	1454
Db	4351	CGACACCTGAGCTACCCCTCGACACAGCGCTCTGAAGAGAGGGTTTCATCACCCAGGCG	4410
QY	1455	ThrProLeuIyTyrAspThrGlyAlaSerThrThrGlySerIlyblyHisAspValIly	1474
Db	4411	ACCCCACTCAAGTACGACTCTGGGGCACCCCTCCACTGGCACCAAGAAACAGAGTGGC	4470
QY	1475	SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla	1494
Db	4471	TTCATCATGGAGCGCCCGCGGCTTCTCTCGCTCCGACCCGCTGGACATATAGCT	4530
QY	1495	AspAlaArgAlaLeuGluArgAlaCysTyrGlnIyIleuSerLeuIlySerArgProGlyThr	1514
Db	4531	GACGCGCGGACCTGGAGCGTGGCTGTATGAAGAGTGTGAAGAGCGGCTCAGGAGCC	4590
QY	1515	AlaSerSerSerGlyIySerIleAlaArgGlyIleProValIleValProGlnIleGly	1534
Db	4591	AGCATGTGTGAGGGGCTTCATACACATGTGGGCTCCAGTGTGTGCTCCGAACTGGGC	4650

QY 1535 LysProArgGlnSerProLeuThrTYrGluAspHisGlyValaProhealaglYHisleu 1554
DB 4651 AAGCCAGCGGAAAGCCACTGACTTGAAGACACGCGGACCCCTTACACAGTCACTG 4710
QY 1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeu 1574
DB 4711 CACGGGGCTCCCTGTGACACGAGGAGCCACGCGCTTACGAGGAGGACGCTC 4770
QY 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlnIleAlaLys 1594
DB 4771 CATTCAGAGAGCGCTCCAGAGACCGGAACTGACATCTACACCCCGGAGATCGCCAG 4830
QY 1595 SerProHisSerThrValProGlnHisSerProHisProIleSerProTYrGlnHisleu 1614
DB 4831 TCCCCACAGACGACTGTGCCCGGACACACCTCACCCCATCTCCCTTGTAGACACTTG 4890
QY 1615 LeuArgGlyValSerGlyValaSerGlyValaSerGlyValaSerGlyValaSerGlyVala 1634
DB 4891 CTCGGGGGGGTGACTGTGTGACCTGTACCTGTGACATCCCATTTGGCTTTGACCCC 4950
QY 1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIleTYrTYrLeuProArg 1653
DB 4951 ACCTCATACCCCGAGGAGTCCCTCTGGAAGCAGACGCGGACCTTACCTGCCCCG 5010
QY 1654 HisleuAlaProAsnProThrTYrProHisleuTYrProProTYrleuIleArgGlyTYr 1673
DB 5011 CACTTGGCCCCCAGCCCACTACACACCTGTACCTTACCTTACCTGCGGCTAC 5070
QY 1674 ProAspThrAlaAlaLeuGlnLeuAsnArgGlnThrIleIleAsnAspTYrIleThrSerGln 1693
DB 5071 CCTGACACGCGCGCTGTGAGAAACCGCAGACCATATCATATCATCATCTGCGAG 5130
QY 1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
DB 5131 CAGATGCACACACACGCTGCTCCGCCATGCGCCAGCGTGTGACATCTGAGGGGTCTG 5190
QY 1714 SerProArgGlySerSerLeuAlaLeuAsnTYrAlaIleGlyProArgGlyIleIleAsp 1733
DB 5191 TCACCGCGAGAGTCCCTGCTGCGCTCATTTATTCGCTGCGCCAGAGGCAATTATTCAG 5250
QY 1734 LeuSerGlnValProHisleuProValLeuValProProThrProGlyThrProAlaThr 1753
DB 5251 CTGTCCCAAGTCCACACCTGCGCTGTGTGCGCACCAAGCGCAGGACCCCTGCGAC 5310
QY 1754 AlaMetAspArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
DB 5311 GCCATGCAGCCGCTTGCCTTACCTCCCACTGCGCCCCCACCCTTCAGCGCGCCCACT 5370
QY 1774 SerSerProLeuSerProGlyGlyProThrHisleuThrLysProThrThrThrSerSer 1793
DB 5371 AGCTCACCGCTGTCCCAAGAGGCCCCCACTCACCTAGCTAAACCACTGCCACATCTTCA 5430
QY 1794 SerGlyValArgGlyValaAspArgArgArgGlyValaAspArgGlyValaLysSer 1813
DB 5431 TCGGAGCGGGAACGGGAAAGTAGCGGGAACGAGAC-----AAGTCC 5472
QY 1814 IleLeuThrSerThrThrThrValaGlnHisAlaProIleTrpArgProGlyThrGlnGln 1833
DB 5473 ATCTCTACGCTTACCATCTACATGAGCATGCACTTGTAGAGACTGTGACGAGCAG 5532
QY 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
DB 5533 AGCAGCGGGGCT-----GGGGGAGAGCGCGCCCGCTCC 5568
QY 1854 HisSerHisAlaHisGlnHisSerProIleSerProArgTYrGlnAspAlaLeuGlnGln 1873
DB 5569 CACACC-----CACAGACACTCGCCCATCTCCCGGAGCAGGAGACCCCTTGCAGCAG 5622
QY 1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValaGluProSer 1893
DB 5633 AGCCCAAGTGTCTGCACAAACGAGCAAGAAAGGGGTGTCACTCTCGTGAACCCGCG 5682
QY 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIleThrPhe 1913

DB 5683 ACCGCCAGGCTCTGAGGTCCACCTCACCTCTCCCTGTGCGCCGACGTGCACATTC 5742
QY 1914 ProProAlaThrHisGlyProLeuGlyGlyTYrThrLeuAspGlyValaTYrProThrLeuMet 1933
DB 5743 CCACCTGCACCCACAGTCCCACTGTGGGACCCCTTGAAGGGGTCTTACCTTACCTCATG 5802
QY 1934 GluProValLeuLeuProLysGlyValaProArgValaAlaArgProGluArgProArgAla 1953
DB 5803 GAGCCGCTCTGTATCCCAAGAGACCTCTCGGGTGGCCCGGCGGAGCGGCCCCGTGTG 5862
QY 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
DB 5863 GACGTTGGCCATGCTCTTCCCTCACCAAAACCCCGCGCG-----GAGCCGCGCTCC 5913
QY 1974 SerProSerGlyLysSerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
DB 5914 TCACCCAGAGAGCTCCAGACCCGATCCCTTACGACACCCCGCTTCAGCCACAGACC 5973
QY 1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAla 2013
DB 5974 ATGGCCGACCCCGAGCAAGAGCCTTGCACCCCACTGACATGCCAGTCCGAGCCGCGGG 6033
QY 2014 ProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPheSerIle 2033
DB 6034 ---CCCACTCGGCTCAGATCTGACACGAGAAAGACTCAAGGTAAACCTTTCCATC 6090
QY 2034 GlnGlnLeuGlnLeuArgSerLeuGlyTYrHis---GlySerSerTYrSerProGlyGly 2052
DB 6091 CAGGAATTGAATCTCGTCTCTGTGGTTACACAGTGGAGCTGGCTTACAGCCCGGATGG 6150
QY 2053 ValGluProValSerProValSerProSerProSerLeuThrHisAspLysGlyLeuProLys 2072
DB 6151 GTGAGGCCATACAGCCCGGTGAGCTCCCACTGTACCCACAGCAAGAGGGCTCTCCAA 6210
QY 2073 HisleuGlnGlnLeuAspLysSerHisleuGlnGlyGlnLeuArgProLysGlnProGly 2092
DB 6211 CCTGTGAAGAGACTGAGAAAGAGCACTTGAAAGGAGACTGCGGACAAAGCAGCAGGC 6270
QY 2093 ProValLysleuGlyGlyGlnAlaHisleuProHisleuAspProLeuProGlyLysSer 2112
DB 6271 CCCATGAGCTAGCGCGGAGCTGCGCATCTCCACATCTCGGCGCTGCGCGAGAG 6330
QY 2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValaLysGlyHisGlnArgVal 2132
DB 6331 CAGCCCTATTCAGCCCACTCTCCAGACTGCCCCAGGATCAAGAGTCAACAGAGGGTG 6390
QY 2133 ValThrLeuAlaGlnHisIleSerGlyValaIleThrGlnAspTYrThrArgHisIlePro 2152
DB 6391 GTACCCCTGGCTCAGACATCAGCAGAGTCATTACGAGAGACTACACCCGCGACACCG 6450
QY 2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTYrSerPheProGlyValaSerGlyPro 2172
DB 6451 CAGCAGCTCAGTGGCCCCCTTCCGCGCTCTACTCTCTTCCGAGACAGCTGCCCC 6510
QY 2173 ValLeuAspLeuAlaArgProProSerAspLeuTYrLeuProProProAspHisGlyAla 2192
DB 6511 GTGCTGATCTTCCGCGCCCACTGACCTTACCTTCCACCCCGGACCATGAGCACC 6570
QY 2193 ProAlaArgGlySerProHisSerGlnGlyLysValArgSerProGluProAsnLysThr 2212
DB 6571 CCAGCCCGGGATCCCCCACTGAAAGGGGCAAAAGGTCCCGGAGAACCCAGCAAAACA 6630
QY 2213 SerValLeuGlyGlyGlyGlnAspArgLysIleGluProValaSerProProGlnGlyMetThr 2232
DB 6631 TCGGTCTCGGAGAGAGTAGGATCCATTGAGCTGTGTCCCAACAGAGGAGCATGACT 6690
QY 2233 GluProGlyHisSerArgSerAlaValTYrProLeuLeuTYrArgAspGlyGlnGlnThr 2252
DB 6691 GAGCCAGAGAGACTGTGAGAGCTGTGTACCATGCTGTATCGAGACGGGAGAACGGGC 6750
QY 2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272


```
Db 6751 GAGCCC--AGATGGGCTTAAGTCTCCAGGCAACACGACGACCGCCACTTCTTC 6807
Oy 2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLysLeuLys 2292
Db 6808 AGTAAGCTGACTGAGAGCACTCCGCCATGTGTGAAGTCAAGAGACAGAGATCAACAAG 6867
Oy 2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThrGlu 2312
Db 6868 AAACCTCAACCCCAACCCGAGACGAGCAAGATACATATTATGGCCAGCTCGGACGGA 6927
Oy 2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 6928 ATCTTCAACATGCGCCCATCATCTGAGCAGAGCCCTTATGACTGTAGAAGCCAGCGGTG 6987
Oy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGlnAlaIleLeuArgLysAlaLeuMetGly 2352
Db 6988 CAAAGAACCGCCAGCAACCAATGGGGCTAGAGGCCATTATTAGAAAGCACTCATGGGT 7047
Oy 2353 LysTyrAspGlnTTPGluGluSerProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db 7048 AAATATGATCAGTGGAGAGAGCCCGCGCTCGGCGCAATGCTTTAAACCTCTGAAT 7107
Oy 2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 7108 GCCAGCGCCAGCTGCGCCGTGCTGTATGCCATACCACTGTAGCGAGCGAGAGTAC 7167
Oy 2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Db 7168 CACGACCTCACCCTCCCGAGGTGGAGGTGGAAAGCCAAAGTCTTGGCAGACTTACGAC 7227
Oy 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 7228 CGAAAGCCCAAGTCCGACAGACCAAGCCCTAGCGTCCGAGAGCCGACCCCTTCTCTCC 7287
Oy 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTTPGlu 2451
Db 7288 TCAGTACACTCAGAGGGGAGTGCATGCGCGAACAACCACTCAACACCGTGTGGAG 7347
Oy 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLysMetArgLeu 2471
Db 7348 GACGGCCCTCATCTGCGAGGGTCCAGCCATTCCCTTACACCTTTGATTATGAGGCTA 7407
Oy 2472 GlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db 7408 CAGGAGAGTGTATGATGCTCCCGCCCACTGGCTTGGCGAGCGGCGCCCTA 7467
Oy 2492 AlaGlyProHisAlaIleTTPAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
Db 7468 GCTGTGCTCCACACGCTCGGATGAGAGCCCAAGCCACTGCTGTTCACAGTATGAG 7527
Oy 2512 ThrLeuSerAspSerGlu 2517
Db 7528 ACACCTCTCGGACAGCGAG 7545

RESULT 13
US-10-087-192-650
; Sequence 650, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 650
```

```
; LENGTH: 7534
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-650

Alignment Scores:
Pred. No.: 0
Score: 9691.50
Percent Similarity: 87.70%
Best Local Similarity: 84.06%
Query Match: 73.34%
DB: 14 Gaps: 28

US-09-522-753-5 (1-2517) x US-10-087-192-650 (1-7534)
Oy 295 LysGlnLysPheCysGlnArgTyrAspGlnLeuMetGlnAlaLeuGluLysValGlu 314
Db 2 GAACAGCCCTTCTGCCAGCCCTATGACCACTCATGAGGGGTGGAGAAAGATAGAG 61
Oy 315 ArgIleGluAsnAsnProArgArgArgAlaLysGlnSerLysValArgGluTyrTyrGlu 334
Db 62 CGCATAGAGAACAAATCCGGAAGAGGGCCAAAGAGAGCAAGTGAAGAGTACTACAG 121
Oy 335 LysGlnPheProGluIleArgLysGlnArgGluLeuGlnArgGluArgMetGlnSerArgVal 354
Db 122 AAACAGTCCCGAGATCCGACAGCGGAGCTGCGAGGCGCATCAGAGCGAGGTG 181
Oy 355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
Db 182 GGCACGCTGGCAGTGGGCTCTCCATGTCGGCTGCCCGCAGTGAAGCAAGAGTTTCTGAG 241
Oy 375 IleIleAspGlyLeuSerGluGlnGlnLeuGlnLysGlnMetArgGlnLeuAlaVal 394
Db 242 ATCATTTAGTGGCTGTGTGACAGAGAACTGGAAGAGAGATGCCAGCTGGCCGTG 301
Oy 395 IleProProMetLeuTyrAspAlaAspGlnGlnArgGluLysPheIleAsnMetArgGly 414
Db 302 ATCCCGCCCATGTTGTATGACGCGGACGACGAGATCAAGTTCATCAACATGATGGA 361
Oy 415 LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTyrSerGlu 434
Db 362 CTGATGATGACCCCATAGATAGGTCTACAAAGACCTCAGGTTTACCAATGTGGAGCGAG 421
Oy 435 GlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIle 454
Db 442 CAGAGAGAGACACCTTCGTAAGAGTTTATGACGACCCCTAAGAACTTTGGCCCTGATT 481
Oy 455 AlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrThrLeuThrLys 474
Db 482 GCCTCATTTCTGAGAGAGAAAGCGGTGCTGAGTGTCTTATTACTTACCTGACCAAG 541
Oy 475 LysAsnGluAsnTyrLysSerLysValArgArgSerTyrArgArgArgGlyLysSerGln 494
Db 542 AAGAAATGAATTTACAAAGACTTGTGAGCGGAGACTTCCGCGCGGTGGCAAGAGCCAG 601
Oy 495 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 602 CAGCAGCAGCAGCAGCAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 649
Oy 515 SerSerGlnGluGluLysAspGluLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 534
Db 650 AGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 709
Oy 535 LysProGluValAlaGluAsnAspLysGluAsnLeuLeuLysGlnLysThrAspAspThrSer 554
Db 710 AAGCAGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769
Oy 555 GlyGluAspAsnAspGluLysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSer 574
Db 770 GCGGAGAGCAACGATGAGAAAGAGCGCTGCGCTCAAAAGCGCGCAAACTCCCAACAGCG 829
Oy 575 GlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGlu 594
```

Dh 830 CAAGGCCCGCAAGGCCCTGATCAAGCGCTCATGGCCAAAGAGCCCAATGAGAG 889
Qy 595 AlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAangLysSerArg 614
Dh 890 ACAGCCACCCCAAGCAAAAGTTCAAGAGCTGGCTTCATGAGAGTAAAGAGAGTTCTCCG 949
Qy 615 TrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGlnHisGlyArgAspTrp 634
Dh 950 TGACATGAGAGAGATGAGACAGCAAGAAAGGCTCTCGAACAATGGAGAGAACTGG 1009
Qy 635 SerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAspPhePhe 654
Dh 1010 TCGAGCATTTCCCGCATGAGTGGCTCCAAAGCCGTGTCCAGGTGAAGAACTTCTACTTC 1069
Qy 655 AsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGlu 674
Dh 1070 AACTACAGAAAGAGCGCAACCTGAGCAAAATCTTACAGACCAAGCTAAAGATGGAG 1129
Qy 675 LysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluAla 694
Dh 1130 AAGGAGAGGAACCTCGAGGAAGAAAGAACCCCACTGGCGGAGCGAGAGACA 1189
Qy 695 AlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGlu 714
Dh 1190 GCTTCCCATCTCCCGCTGAGGACGAAGATGGAAGCATCAGCGCAAGTCCCAATGAG 1249
Qy 715 GluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArg--- 723
Dh 1250 GAAGAGCTGGCGGAGAGAGCAAGCTTCAAGGCTCTGGAAATGAGTTCCAGAGTT 1309
Qy 734 GlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSer 753
Dh 1310 GGGAGTGCAGTGGCCAGCTGCTCAACAAGCTCTGATACAGAGTGTCCATCC 1369
Qy 754 ProHisThrGluAlaAlaLysAspThrGlyLysAsnGlyProLysProProAlaThr--- 772
Dh 1370 CCGCTTCAAGAACCGCAAGAGCACT-----GGGCTTAAACCCATGCGACCTGAA 1420
Qy 773 ---LeuGluValAspGlyProProGluProProThrProProArgArgThrSerArg 791
Dh 1421 GCATTCCCGGCTGCCACCGCACTGTCTCT-----CTTCAAGAAACCGGCGCA 1474
Qy 792 AlaProIleGluProThrProAlaSerGluAlaThrGlyValaProThrProProAla 811
Dh 1475 GCCCTGCTGAGGCTTCCCAAGTCCCGTANAGCAGAGGCCACCAATCCAGAGCTTCC 1534
Qy 812 ProProSerProSerAlaProProProValValProLysGluGluLysGluGluThr 831
Dh 1535 ---CCATCACTGCGCCAGCCCGGCTACTGTGACAAAGATGAACAAGAACCCCGGCT 1591
Qy 832 AlaAlaAlaProProValGluGluGluGluGluGluGluProProAlaAlaGluGluLeu 851
Dh 1592 GCTCCAGCTTCCCAAGCAAGATGCCAAGAGCAAGAGCTGAGGCCGAGAG--- 1645
Qy 852 AlaValAspThrGlyLysAlaGluPro-----Val 862
Dh 1646 ---ATCATGTGGGAAGCAAGAGAGCCGAGGCTGAGAGGCCCGGAGAGTGA 1702
Qy 866 LysSerGluCysThrGluGluAlaGluGluGlyProAla---LysGlyLysAspAlaGlu 881
Dh 1703 AAGAGGACCAAGAGAGAGCCGAGGAAGCTGGAAGACAAACCCAAAGGCGACAGAG 1762
Qy 882 AlaAlaGluAlaThrAlaGluGluAlaLeuLysAlaGluLysLysGluGlyGlySerGly 901
Dh 1763 GCCATTGAAACTGTGTCTAGAGCACACTTAAGTGGAG-----GAGGCTGTACCAAG 1816
Qy 902 ArgAlaThrThr---AlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Dh 1817 GCAGCTGTACCAAGGCTTCAAGTCAAGGAGCCACAGGACAGTCACTCAAGTGCACAC 1876
Qy 921 CysSerAlaAspGluValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro 940
Dh 1877 TGCAGTGCAGATGAGGTGACGAACCGAAGAGGTGAACAAGGCGGCTGTCTACCA 1936

Qy 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaAlaSerProGlnLysPro 960
Dh 1937 AGGCCCAAGCTCTCTACCCCGGCTGAGAGATCCCGGCGCAATCTGCCCCCAAGCCG 1996
Qy 961 LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLys 980
Dh 1997 CTGAGCTTGAAGCAGTGAAGCAGCAGCAGCGCCATCCCGCTATC---GTCACCAAG 2053
Qy 981 ValHisGluProProAlaGluGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Dh 2054 GTCATGAGCCCCCGGAGAGACACAGTACCCCAAGGACAGTTCCCGCTGCTTCCA 2113
Qy 1001 ProProGlnAsnLeuGlnProGlnLysAspAlaProGlnGlnProGlnLysSerSerProArg 1020
Dh 2114 CCAAGCAGCACTTACAGCCAGAGGAGTGTCTGTACAGATGTGGAGAGATCCAGT 2173
Qy 1021 GlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Dh 2174 GGCAGTCCCGCAGCCAGTGCCTCGCCGAGAAAGAGCGACGAACCCGATCTTT 2233
Qy 1034 ---AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGly 1052
Dh 2234 CCGGCTTCCCAACTGAGGCGCCAAAGCTACAGACTGAGCCCGCCAGCTGTATCGGCG 2293
Qy 1053 LeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro 1072
Dh 2294 CTGCGCTTCCCAATCCCTCCAGGAGGTATCAAGCTTCCCAAGCGCGCTGAGCCC 2353
Qy 1073 SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla 1092
Dh 2354 TCGCGCTTCTCTACACACCCCGGTCACCGCTGCTGTGGGCTTCCAGATAGGCC 2413
Qy 1093 ArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAla 1112
Dh 2414 CCGCGCGCTGTCAGTCCCGCC---ATCTTAAACCCCAACCCCTCATCTCTCGCC 2470
Qy 1113 LysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnLysMetSerValGln 1132
Dh 2471 AAGCATCCCGGCTACTTGAAGGAGCTGGGAGCCATCTTCCAGGAGATGTCAAGTCCAG 2530
Qy 1133 LeuHisValProTyrSerGlnHisAlaLysAlaProValGlyProValThrMetGlyLeu 1152
Dh 2531 CTTCGATGCTCACTCAGAGCATGCAAGGCCCATGAGGCTCTCAACATGGGGCTG 2590
Qy 1153 ProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeu 1172
Dh 2591 CCCCTTGGCGTGAACCTTAAGAAAGCTG----- 2617
Qy 1173 SerProArgGlyGlnAlaGlyProProGlnSerLeuGlyValProThrAlaGlnGluAla 1192
Dh 2617 ----- 2617
Qy 1193 SerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIle 1212
Dh 2618 -----GGAGACACACTGGCTCGGACACAGAGGAAGCATTCACCAAGGGGCTC 2665
Qy 1213 ProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGly 1232
Dh 2666 CCCAGTACCAGGCTGACAGCGGCC-----AGCTACAGAGGCTCTATCACACCGCG 2719
Qy 1233 ThrProAlaAspValLeuThrLysGlyThrIleThrArgIleIleGlyGluAspSerPro 1252
Dh 2720 AGCCCGCAGAGCTCTTACAAAGGTACCATACAGAGATGTCTGGTAGGACAGCCA 2779
Qy 1253 SerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGly 1272
Dh 2780 AGTCCGCTTGACCGGCAAGAGGACACCTCGCCCAAGGCGCATTCATATAGAGGGC 2839
Qy 1273 LysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGlu 1292
Dh 2840 AAGAAAGGCGACGTCTTATCTATGAAGGTGATGTCTGTGTACAGAGTCTTAAAGAG 2899

QY 1293 AspglyarGserSerSerGlyProProhiIsGluThrAlaAlaProLysArgThrTyrsP 1312
DB 2900 GATGAGAAAGACAGCTCGGCGCCACCACCTAGAGACTGGCGCCCTTAAGACCACTATGAC 2959
QY 1313 MetMetGluGlyArgValAGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGly 1332
DB 2960 ATGATGAGAGGCGCGTGTAGGACAGACTGTACACTCAGCAGACATAGAGGACTCATGGGC 3019
QY 1333 ArgAlaIleProProGluArgHisSerProhiShiSleuLysGluGlnHisIleArg 1352
DB 3020 CGCGGCATC-----CCTGACAGACAGACCC---CACCTCAAGACAGCATCACTCCGA 3073
QY 1353 GlySerIleThrGlnGlyIleProArgSerTyrsValAGluAGluAspTyrsLeuArg 1372
DB 3074 GGCCTCATACCGCAAGGACCTCCGAGGTCTATGTGAGGCGCAGAGAGACTTACGG 3133
QY 1373 ArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArgAsp 1392
DB 3134 CGGAGGCCAAGCTCTTGAAAGGAGAGAGACACCAACCCCAACCACTCGGGAGC 3193
QY 1393 LeuThrGluAlaTyrsThrGln-----AlaLeuGlyProLeuLysLeuLysPro 1409
DB 3194 CTGACTGAGACTTACAAAGCCCGCCCTGACCCCTGCGGTCCCTGAAAGCTGAAGCGG 3253
QY 1410 AlaHisGluGlyLeuValAlaThrValIysGluAGlyArgSerIleHisGluIlePro 1429
DB 3254 ACTCAGAGGGGTGTGTAGCAACTGTGAAGAGGCGGCGGCTTATTCATAGATCCCG 3313
QY 1430 ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
DB 3314 AGAGAGAGACTCGCGCGCACACTGAGCTAACCCCTGGCACCAAGGCTCTGAAGAGGGGT 3373
QY 1450 SerIleThrGlnGlyThrProLeuLysTyrsAspThrGlyAlaSerThrThrGlySerLys 1469
DB 3374 TCCATCACCCAGGGGACCCCACTCAAGTACGACTCGGGGACCCCTCACTGGCAACAG 3433
QY 1470 LysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPro 1489
DB 3434 AAACAGAGGTGGCTCCATCATCGGACAGCCCGGCGGCTTTCCCTGCGCCGACCCG 3493
QY 1490 LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrsGluGluSerLeuLys 1509
DB 3494 CTGGACATATATGGCTGATCCCGGGGCACTGGAGCGTCCGCTATGAAAGAGACTGAAG 3553
QY 1510 SerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIle 1529
DB 3554 ACCCGGTCAAGGACAGAGGTGTGCAAGGGGGCTCCATCACACGTGGGGCTCCAGTCTC 3613
QY 1530 ValProGluLeuGlyLysProArgGlnSerProLeuThrTyrsGluAspHisGlyAlaPro 1549
DB 3614 GTGCTGAACTGGGCAAGCCACGGCAAGCCCACTGACTTACAAAGACACGGGGGACCC 3673
QY 1550 PheAlaGlyHisLeuProArgGlySerProValIleMetArgGluProThrProArgLeu 1569
DB 3674 TTACACAGTCACTGCGCACGTGGCTCCCTGTGACACAGAGGAGCCAGCACCGCTT 3733
QY 1570 GlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro 1589
DB 3734 CAGGAAGGACGCTCTCATCCAGCAAGCGCTCCAGAGACCGGAAGCTGACATCTACACC 3793
QY 1590 ArgGluIleIleAlaLysSerProHisSerThrValProGluHisIleAspProIleSer 1609
DB 3794 CGGAGATTCGCCAAGTCCCAACAGACATGTGCGCAGACACCACTTCACTCCCATCTCC 3853
QY 1610 ProTyrsGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrsArgSerHisIlePro 1629
DB 3854 CCTATGAGACACTTCTCCCGGGCGTACGTGTGGAACCTGTACCGTGTACATCCCA 3913
QY 1630 LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAla 1648
DB 3914 TTGGCCTTTGACCCCACTCATATCCCGAGGAGATCTCTGTGAAGAGACAGCGCGACCC 3973
QY 1649 TyrTyrsLeuProArgHisLeuAlaProAsnProThrTyrsProHisLeuTyrsProProTyrs 1668

DB 3974 TACTACCTGCCCCGCGACTTGCGCCCGCCAGCCCACTTACCAACCTGTATCCACTTAC 4033
QY 1669 LeuIleArgGlyTyrsProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
DB 4034 CTCAATCGCGGCTACCTTGACAGCGGCGCTTGAGAAACCCCAACACATCATCATATAC 4093
QY 1689 TyrIleThrSerGlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
DB 4094 TACATACCTCGAGACATATCAACACACCTGCTCCGCAATGGCCACAGTGTCTGAC 4153
QY 1709 MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrsAlaGlyPro 1728
DB 4154 ATGCTGAAGGGGTGTGTCAACCGAGATCTTCGCGGCCCTCAATATAGCGGTGGCCA 4213
QY 1729 ArgGlyIleIleAspLeuSerGlnValProHisLeuProValIleValIleProThrPro 1748
DB 4214 AGAGGCAATTATGACTTCTCCAAAGTGCACACCTGCGCTGTGTGTGCACCAAGCCA 4273
QY 1749 GlyThrProAlaThrAlaMetAspArgLeuAlaTyrsLeuProThrAlaProGlnProPhe 1768
DB 4274 GGACACCTTGCCACCGGCATGAGCCGCTTGCTTACTCTCCACTGGCGCCCACTTCC 4333
QY 1769 SerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisSleuThrTyrsPro 1788
DB 4334 AGACACCGCCACAGTAGTCAACCGCTGTCCAGAGAGGCCCACTCACCTTAAACCA 4393
QY 1789 ThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArg 1808
DB 4394 ACTGCCACATTTCTCATCGAGCGGAGAGGAGAACTGAGCGGAAAGAGAC----- 4444
QY 1809 GluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleTyrArg 1828
DB 4445 -----AATCATCTCTCAACGTCTACACTTACAGTGAACATGACCCCATCTGAGA 4495
QY 1829 ProGlyThrGlnGlnSerSerGlySerSerGlyGlyGlyGlyGlySer 1848
DB 4496 CTGTGTACGAGACAGACGACGCGGGCT-----GGGGGCACG 4531
QY 1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
DB 4532 AGCCGCGCCGCTCCACAC-----CACAGCACTGCCCATCTCCCGGGAGCCAG 4585
QY 1869 AspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThr 1888
DB 4586 GAGCGCTTGACAGAGAGGCCAGTGTGTGCAACAACAGACATGAAGGGCGTGTACCC 4645
QY 1889 Ala-ValGluProSerLysProThrValLeuArgSerThrSerThrSerProValArg 1908
DB 4646 TCCCGTGAACCCGGGACGCCACCGGTCTGTAGGTCCACCTTCCACTTTCGCTGTCCG 4705
QY 1908 GProAlaAlaIleThrPheProProAlaIleHisCysProLeuGlyGlyThrLeuAspGlyVal 1928
DB 4706 CCAGGTGCACATTTCCACCTGCGCACCACTGCGCACTTGTGTGCACTTGAAGGGGT 4765
QY 1928 TyrProThrLeuMetGluProValIleLeuLeuProLysGluAlaProArgValAlaArgPro 1948
DB 4766 CTACCTTACCTCATGAGACCGGTCTCTTATCCCAAGAGAACTTCGGGTGCCCGGCC 4825
QY 1948 OGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGly 1968
DB 4826 CGAGCGGCGCCGTGTGAGAGCTGGCCATGCTTCTTCCACCAACCCCGCGCGG----- 4880
QY 1968 yLeuGluProAlaSerSerProSerTyrsGlySerGluProArgProLeuValIleProProVa 1988
DB 4881 ----GAGCCGCTTCTTCAACCAAGAAAGCTCGAGCCCCGATCCCTTACACCCCCAG 4936
QY 1988 lSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisIleAlaSe 2008
DB 4937 CTCACGACCAACAGCATGCGCGGCAACCCAGAAAGAGCTTGCACCCACCATGCGCAG 4996
QY 2008 rProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSe 2028

Db 4997 TCCGAGCCCGCGCG---CCCACTCGGCTCGATCTCGACCGAGAAAAGACTCAAG 5053
Qy 2028 rlyserProPheSer11egIngluLeuGluLeuArgSerLeuGlyTyris--glyserse 2047
Db 5054 TAAACCTTTTCCATCCAGAAATTGGAATCTCGTTCTCTGGGTTTACACAGTGAAGCTGG 5113
Qy 2047 rTyserProGluGlyValGluProValSerProValSerProSerLeuThrHis 2067
Db 5114 CTACAGCCCCGATGAGGTGAGCCATCAGCCCGGTAGCTCCCGACCTGACCCAGCA 5173
Qy 2067 pLyserGlyLeuProLyserHisLeuGluGluLeuAspLyserHisLeuGluGlyLeuLeu 2087
Db 5174 CAAGGGGCTCTCCAAACCTCTGGAAGAGCTGAAGAAAGCCACTGGAAGGGAGCTGG 5233
Qy 2087 gProLyserGluProGlyProValLyserLeuGlyGlyValAlaHisLeuProHisLeu 2107
Db 5234 GCACAGACCCAGGCCCCCATGAAAGCTCAGCGGGAAGGCTGCCATCTCCACATCTGG 5293
Qy 2107 gProLeuProGluSerGluProSerSerSerProLeuLeuGluThrAlaProGlyVal 2127
Db 5294 GCCACTGCCGAGAGCCAGCCCTCATCAGCCCACTCTCCAGACTGCCCGACGCACTCA 5353
Qy 2127 gGlyHisLeuArgValValThrLeuAlaGluHisAlaLeuSerGluVal11ethrGlu 2147
Db 5354 AGGTCAACCAAGAGGTGTCTCACTGCTCAGACATCAGGAGGTCTTACGAGACTA 5413
Qy 2147 rThraGHisHisProGluGluLeuSerAlaProLeuProAlaProLeuTyserPhe 2167
Db 5414 CAGCGCCACACCCGACAGACTCACTGCGCCCTCCGCGCTCTCACTCTCTTCC 5473
Qy 2167 ogLyAlaSerCyProValLeuAspLeuArgArgProProSerAspLeuTyserPro 2187
Db 5474 CGAGAGCAGCTGCTCTCTCTGATCTTCCGCCGCCACCGAGACTTACCTCCAC 5533
Qy 2187 oProAspHisLeuValAlaProAlaArgLyserProHisLeuSerGlyGlyTyser 2207
Db 5534 CCCCAGCAAGGAGACCCAGCCCGGGATCCCCCAAGAAAGGGGGCAAAAGTCCCC 5593
Qy 2207 ogLProAspLeuTherSerValLeuGlyGlyGlyAspGly11egLProValSer 2227
Db 5594 AGAACCCAGAAACATCGTCTCTGGGAGAGTGAAGATGCCATTTAGCTGTCTCCC 5653
Qy 2227 oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyserLeuLeuTy 2247
Db 5654 ACCAGAGGGGATGATCAGAGCCAGACATGCTCGAAGGCTGTATCCCATGCTATCG 5713
Qy 2247 gAspGlyGlyGluThrGluProSerArgMetGlySerTyserProGlyValanTher 2267
Db 5714 AGACGGGGAAACAGGGGAGGCC---AGATGGGCTCTTAAGTCTCAGGCAACCA 5770
Qy 2267 nProProAlaPhePheSerTyserLeuThrgLuseArganSerAlaMetValLyserTy 2287
Db 5771 GCCGCAAGCTTTCTTCACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5830
Qy 2287 gGluGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 2307
Db 5831 GCGAGGATCAACAAGAACTCAACCCCAACCGGAAGACAGACCAAGAAATATGG 5890
Qy 2307 rGluProGlyThrGluLeuPheAspMetProAla11ethrGlyTyserLeuMetTh 2327
Db 5891 CCAGCTCGGAGCGGAATCTTCAACATGCCCCGACATCAGGAGCGGCTTATGACTG 5950
Qy 2327 rArgSerGluAlaValGluGluLeuHisLeuThraMetGlyLeuGluAla11eLeu 2347
Db 5951 TGAAGCCAGGCGCTGCAAGAACAGCCAGACCAATAGGGCTGAGGCCATTTATG 6010
Qy 2347 gLyValAlaLeuMetGlyTyserTyserArgLntPrGluGluSerProProLeuSerAla 2367
Db 6011 AAAGGACATCATGTAATATATGATCATGTGAAGAGCCCCCGCTGGCCCATG 6070
Qy 2367 aPheAspProLeuAspAlaSerAlaSerLeuPro--AlaAlaMetPro11ethrAla 2386
Db 6071 TTTTAACTCTGTAATGCGAGCCAGTGTGCCGCTGTCTATGCCCCATTAACACTGC 6130

Qy 2386 aAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyTyserAlaTyserVal 2406
Db 6131 TGAAGAGGAGTGAACCAACCACTCACTGCGAGAGTGAAGTGAAGAAAGCTCTC 6190
Qy 2406 rGlyArgProSerSerArgLyAlaTyserSerProAlaProGlyLeuAlaSerGlyAs 2426
Db 6191 TGGCAGACCTAGAGCGGAAAGCAAGTGCAGACACCAAGGCTTACCTCGGAGACG 6250
Qy 2426 gProProSerValSerSerValHisSerGlyGlyAspCyAspArgArgThrProLeu 2446
Db 6251 ACCCTTCTGTCTCTCACTACATCAGAGGGGAGTGAATGCGGAACACCACTCAC 6310
Qy 2446 rAspArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyra 2466
Db 6311 CAACCGTGTGGAGAGACCGGCCCTCATCTGAGAGGTCAACCCATTCCTCAACACC 6370
Qy 2466 oLeu11eMetArgLeuGluAlaGlyValMetAlaSerProProProGlyLeuProAl 2486
Db 6371 TTTGATTATGAGGCTACAGGAGGATGATGCTCCCGCCCACTGGCTGGCGG 6430
Qy 2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGlyGluProLyserPro 2506
Db 6431 AGGACAGGGGCTCTAGCTGTCTCCACACCGCTGGAGTGAAGAGCCAAAGCACTG 6490
Qy 2506 uCySerGluTyserGluThrLeuSerAspSerGlu 2517
Db 6491 GTCTTCAAGTATGAGACACTCTCGACAGCGAG 6524

RESULT 14
US-10-146-473-14
; Sequence 14, Application US/10146473
; Publication No. US2003010888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146, 473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291, 150
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-146-473-14

Alignment Scores:
Pred. No.: 7, 66e-264 Length: 2930
Score: 4612.00 Matches: 873
Percent Similarity: 94.38% Conservative: 1
Best Local Similarity: 94.28% Mismatches: 6
Query Match: 34,90% Indels: 46
DB: 16 Gaps: 1

US-09-522-753-5 (1-2517) x US-10-146-473-14 (1-2930)

Qy 1592 l1eAlaLyserProHisSerThrValProGluHisHisProHisProHisSerProTy 1611
Db 123 CTGGCTCAACCCCGACAGCAACCGTCCGAGACACCAACCCCACTTCTGCTTAT 182
Qy 1612 GluHisLeuLeuArgGlyValSerGlyValaAspLeuTyArgSerHis11eProLeuAla 1631
Db 183 GAGCAGCTCTCTGGGGGCTGAGTGGCTGTATGCGAGCCCATTCCTCGGCC 242
Qy 1632 PheAspProTherSer11eProArgGly11eProLeuAspAla11eAlaTyTyserLeu 1651

|||||
Db 243 TTCAGACCCCACTCCATACCCCGCGGCATCCCTCTGAGACGACGCCCTCACTACCTG 302
Qy 1652 ProAaRgH1sLeuAlaProAsnProThrTyrrProH1sLeuTyrrProProTyrrLeu1leAaRg 1671
Db 303 CCCCAGACCTGAGCCCCCAACCCCACTACCCGACCTGATACCCCACTCACTCACTCCG 362
Qy 1672 G1yTyrrProAspThrAla1aleuGluAaAsnArgGlnThr1le1leAsnAspTyrr1leThr 1691
Db 363 GGCTACCCCGACACGCGCGGCGTGGAGAACCGGCACATCATCATCATCATCATCATCAC 422
Qy 1692 SerGlnGlnMetH1sH1sAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAaRg 1711
Db 423 TCGACAGCATATGACACACACGCGGCACCGCATGCGCCAGCGACCTGATATGCTGAAG 482
Qy 1712 G1yLeuSerProArgGluSerSerLeuAla1eAsnTyrrAla1aG1yProArgG1y1le 1731
Db 483 GGCTCTCGCCCCGACAGTCTCTGCGCATCTACCTAGCTGCGAGGTCCCGAGGCATC 542
Qy 1732 1leAspLeuSerGlnValProH1sLeuProVal1leuValProProThrProG1yThrPro 1751
Db 543 ATCGACCTGTCCCAATGGCCACACCTGCGTGTCTGTGCCCCGACACACGACCCCA 602
Qy 1752 AlaThrAlaMetAspArgLeuAlaTyrrLeuProThrAlaProGlnProPheserSerArg 1771
Db 603 GCACCGCATGAGACCGCTGTGCTACTCCCAACCGCGCCCAAGCCCTTCAGACACCGC 662
Qy 1772 H1sSerSerSerProLeuSerProG1yG1yProThrH1sLeuThrTyrrProThrThr 1791
Db 663 CACAGCAGCTCCCACTCTCCCGAGAGGTCCAACACACTTGACAAACACACACACAG 722
Qy 1792 SerSerSerG1yAaRgG1yAaRgAspArgAspArgG1yAaRgAspArgG1yAaRgG1y 1811
Db 723 TCTCGTCCGAGCGGAGCGAGACCGGATCGAGACCGGAGACCGGATCGGAGCGGAA 782
Qy 1812 TyrrSer1leLeuThrSerThrThrValGluH1sAlaPro1leThrArgProG1yThr 1831
Db 783 AAGTCAATCTCACTCACTCAACACAGACGATGAGACAGCACCCATCTGAGACCTGTATA 842
Qy 1832 G1yGlnSerSerG1ySerSerG1ySerSerG1yG1yG1yG1ySerSerSerArgPro 1851
Db 843 GAGCAGACGACGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 902
Qy 1852 AlaSerH1sSerH1sAlaH1sGlnH1sSerPro1leSerProArgThrGlnAspAla1e 1871
Db 903 GCCTCCACTCCATCCACACGACATGCGCATCTCCCTCGACCCACGATGCGCTC 962
Qy 1872 GlnGlnArgProSerVal1leuH1sAsnThrG1yMetLyG1y1le1leThrAlaVal1Glu 1891
Db 963 CAGCAGACGACCGATGTGCTTCAACACAGCATGAGGTATCATCACCGCTGTGAG 1022
Qy 1892 ProSerLyProThrAla1eAsnArgSerThrSerSerSerProAla1eArgProAla1a 1911
Db 1023 CCAAGACGCGCCAGGCTGAGGTCACTCACTCACTCACTCACTCACTCACTCACTGCG 1082
Qy 1912 ThrPheProProAlaThrH1sCyrrProLeuG1yG1yThrLeuAspG1yValTyrrProThr 1931
Db 1083 ACATTCACACCTGCGCACCACTGCGGCGGACCTCGATGGGGGTTCACCTTACC 1142
Qy 1932 LeuMetGluProVal1leuLeuProLyG1yAlaProArgValAlaAaRgProG1yAaRgPro 1951
Db 1143 CTCATGAGACCGGCTTGTGCGCCAGAGAGGCCCGCGGTGCGCGCGCAGAGCGGCC 1202
Qy 1952 ArgAlaAspThrG1yH1sAlaPheLeuAla1yAsProProAla1eArgSerG1yLeuGlnPro 1971
Db 1203 CAGACGACACCGGACATGCTTCTCGCCAGACCCCAAGCGCTTCGCGGCTGGAGCCC 1262
Qy 1972 AlaSerSerProSerLyG1ySerG1yProArgProLeuValProProValSerG1yH1s 1991
Db 1263 GCCTCTCCCGCAGACGAGGCTGAGAGCCCGCGCTAGTGCCTCTGCTGTGCGCAC 1322
Qy 1992 AlaThr1leAlaArgThrProAla1yAsnLeuAlaProH1sH1sAlaSerProAspPro 2011
|||||

Db 1323 GCCACCATCGCCCGACCCCTCCGAAAGAACTTCGCACCTCACACGACGCCCGGACCGC 1382
Qy 2012 ProAlaProProAlaSerAlaSerAspProH1sArgG1yTyrrThrGlnSerLyProPhe 2031
Db 1383 CCGCGCGCACCTGCTCGGCTCGGACCCGACCCGGAAGAAAGACATCAAGTAAACCTTT 1442
Qy 2032 Ser1leGlnGlnLeuG1yLeuAaRgSerLeuG1yTyrrH1sG1ySerSerTyrrSerProGlu 2051
Db 1443 TTCATTCAGGAACGTGAATCCGTTCTGTGGGTTCACAGGACGACGTACAGCCCGAA 1502
Qy 2052 G1yValGluProValSerProValSerSerProSerLeuThrH1sAspLyG1yLeuPro 2071
Db 1503 GGGGTGAGCCCGCTCAAGCCCTTGAGCTCAACCATCTACCCACACAGAGGGGCTCCCC 1562
Qy 2072 TyrrH1sLeuGluGluLeuAspLySerH1sLeuGlnG1yG1yLeuAaRgProLyGlnPro 2091
Db 1563 AAGCACTTGAAGAGCTGACAAAGGCCACTGAGAGGGGAGACTGGGCCCAAGACGCA 1622
Qy 2092 G1yProVal1yLeuG1yG1yG1yAla1aH1sLeuProH1sLeuAaRgProLeuProGlu 2111
Db 1623 GGCCCCGTGAACCTTGCGGGGAGGCGCCCACTCCACACTGCGCGCGCTGCTGAG 1682
Qy 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProG1yValLyG1yH1sGlnArg 2131
Db 1683 AGCCAGCCCTCGTCCAAGCCCGCTGCTCAAGACCGCCCAAGGGGTCAAGGTCAACGCG 1742
Qy 2132 Val1Val1ThrLeuAlaGlnH1s1leSerGluVal1leThrGlnAspTyrrThrArgH1sH1s 2151
Db 1743 GTGTCAACCTTGCCACAGCATCATGATGATCATCATCAACAGACTACACCCGGCACAC 1802
Qy 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrrSerPheProG1yAlaSerCyrr 2171
Db 1803 CCACAGCAGCTCAAGGCACCCCTGCGCCGCCCTTACTCTTCTCTGAGGCGCACTGC 1862
Qy 2172 ProVal1leuAspLeuAaRgRgProProSerAspLeuTyrrLeuProProProAspH1sG1y 2191
Db 1863 CCGTCTCGAAGCTTCGCGCCGCCCAACCATGACCTTACTCTCCCTGAGGCGCACTGC 1922
Qy 2192 AlaProAlaArgG1ySerProH1sSerG1yG1yLyAaRgSerProGluProAsnLy 2211
Db 1923 GCCCGGCGCGTGGCTCCGCCACAGCGAAGGGGCAAGGTCTCCAGAGCCAAACAAG 1982
Qy 2212 ThrSerVal1leuG1yG1yG1yGluAspG1y1leGluProVal1eSerProProGluG1yMet 2231
Db 1983 AGCTGCGCTTGGGTGTGTGAGGACGGTATTTGAACCTGTCTCCACCGGAGGGCAAG 2042
Qy 2232 ThrGluProG1yH1sSerArgSerAlaValTyrrProLeuLeuTyrrArgAspG1yGluGln 2251
Db 2043 ACGAGGCCAGGCGACTCCCGAGGTGTGTACCGGCTGCTTACCGGATGGGGAACAG 2102
Qy 2252 ThrGluProSerArgMetG1ySerLySerProG1yAsnThrSerGlnProProAlaPhe 2271
Db 2103 ACGAGCCACAGAGATGGGCTCCAAGTCTCAAGGAACACACAGCAGCGCGACGCTTC 2162
Qy 2272 PheSerLyLeuThrGluSerAsnSerAlaMetVal1ySerLyLyGlnGln1leAsn 2291
Db 2163 TTCAGAGACTGACCGAGAGCAACTCCGCAATGATCAATCAATCAATCAAGGACGATCAAC 2222
Qy 2292 TyrrLyLeuAsnThrH1sAsnArgAsnGluProGluTyrrAsn1leSerGlnProG1yThr 2311
Db 2223 AAGAACTGTAACACCAACCGGAATGAGCTGAAATCAATCAATCAAGCAGCTGGGAG 2282
Qy 2312 G1y1lePheAsnMetProAla1leThrG1yTyrrG1yLeuMetThrTyrrArgSerGlnAla 2331
Db 2283 GAGATCTTCAATATGCGCGCATACCGGAACAGGCGTTATGACCTATGAAGCCAGGCG 2342
Qy 2332 ValGlnGluH1sAlaSerThrAsnMetG1yLeuGluAla1le1leArgLyVal1leuMet 2351
Db 2343 GTGACGAAACATGACAGCACCAATGCGGTGAGGCCATATTAAGAAAGCACTCAAG 2402
Qy 2352 G1yTyrrTyrrAspGlnThrProGlu1eSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
||||| 2402 ----- 2402

QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGlyGlyGlyGlyValAlaValSerGlyArgProSerSer 2411
Db 2403 -----GGTGGCGGGGAGGCCCAAGGTCTCTGGCAGACCCAGCAGC 2444
QY 2412 ArgValAlaValSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 2445 GGAAGACCAAGTCCCGGCGCCCGGCTGGCATCTGAGACCGGCCACCTCTGTCTCC 2504
QY 2432 SerValHisSerGlyGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrrGlu 2451
Db 2505 TCAGTGCACTCGAGGGAGAGACTGCACCGCGGACCGGCTGCACCAACCGGTGTGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db 2565 GACAGGCGCTCGTCCGAGGTTCCACCGCCATTCCTTCAACACCCCTGATCATCGGCTG 2624
QY 2472 GluAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db 2625 CAGGCGGCTGTCAAGCTTCCACCCGACCGGCGCTCCCGGGGAGGCGGCGCTC 2684
QY 2492 AlaGlyProHisAlaValTrrAspGlyGlyProLysProLeuLeuCysSerGlnTrrGlu 2511
Db 2685 GCTGGGCGCCACACCGCTGGAGAGAGAGCCCAAGCCTGCTCTCTGCACTAGAG 2744
QY 2512 ThrLeuSerAspSerGlu 2517
Db 2745 ACACTCTCCGACAGCGAG 2762

RESULT 15

US-10-174-014-13
Sequence 13, Application US/10174014
Publication No. US20040005292A1
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freiler
APPLICANT: Kenneth W. Doble
TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
FILE REFERENCE: PFS-0012
CURRENT APPLICATION NUMBER: US/10/174,014
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 73
SEQ ID NO 13
LENGTH: 2930
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (456) ... (2765)
US-10-174-014-13

Alignment Scores:

Pred. No.: 7,66e-264 Length: 2930
Score: 4612.00 Matches: 873
Percent Similarity: 94.38% Conservative: 1
Best Local Similarity: 94.28% Mismatches: 6
Query Match: 34,90% Indels: 46
DB: 18 Gaps: 1

US-09-522-753-5 (1-2517) x US-10-174-014-13 (1-2930)

QY 1592 IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyr 1611
Db 123 CTGGGCTCAACCCCGACACACCGTGGCCGAGACACCAACCCCATCTGGCCTAT 182
QY 1612 GluHisLeuLeuArgGlyValSerGlyValAspLeuTrrArgSerHisValLeuAla 1631
Db 183 GAGCACCCTGCTGGGCGGTGAGTGGCTGAGACTGTATGACAGCCACATCCCTGGCC 242

QY 1632 PheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTrrTrrLeu 1651
Db 243 TTGACCCCACTCCATACCCCGGCGATCCCTCTGAGACCGAGCGGCTACTACTG 302
QY 1652 ProArgHisLeuAlaProAsnProThrTrrProHisLeuTrrProTrrTrrLeuIleArg 1671
Db 303 CCCGACACCTGGACCCCAACCCCACTTACCCGACCTGTATACCACTTACTATCCG 362
QY 1672 GlyTrrProPheThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnProTrrIleThr 1691
Db 363 GGTACCCCAACAGCGCGGCTGGAGAACCGGACCATCATCAATGACTATCATACC 422
QY 1692 SerGlnIleMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711
Db 423 TCCGACAGATGACACCAACCGGCGCACCGCATGGCCAGCAGCTGATATGCTAGG 482
QY 1712 GlyLeuSerProArgGlySerSerLeuAlaLeuAsnTrrAlaAlaGlyProArgGlyIle 1731
Db 483 GGCCTCTCGCCCGGAGTCTCGTGGCACTCACTACGCTGGGGTCCCGAGGATC 542
QY 1732 IleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro 1751
Db 543 ATGACCTGTCCCAAGTGCACACTGCTGTGCTGTGCCCCGACACAGGACCCCA 602
QY 1752 AlaThrAlaMetAspArgLeuAlaTrrTrrLeuProThrAlaProGlnProPheSerArg 1771
Db 603 GCCACCGCATGAGACCGCTTGCCTTACTCCACCGGCGCCAGCCCTTACAGACGCC 662
QY 1772 HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThr 1791
Db 663 CACAGAGCTCCCACTCTCCAGAGGTCACAACTTGAACAAACCAACACACAG 722
QY 1792 SerSerSerGluArgGlyValArgAspArgAspArgGluArgAspArgGluArgGlu 1811
Db 723 TCTCTGTCCAGAGGAGAGACCGGATCGAGAGGAGACCGGATTCGGAGCGGAA 782
QY 1812 LysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrrAspProGlyThr 1831
Db 783 AAGTCAATCTTCACGTTCACACACGAGGTGAGACAGCACCTTGTGAACTGTGTACA 842
QY 1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgPro 1851
Db 843 GAGCAGAGCAGCGGACGAGCGGACAGCGGCGGGGTGGGGAGACAGACCGGCC 902
QY 1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871
Db 903 GCCTCCCACTCCATGCCCCACAGCACTGCCCCATCTCCCTGGAGCCAGATGCCCTC 962
QY 1872 GluGlnArgProSerValLeuHisAsnThrGlyMetLeuGlyIleIleThrAlaValGlu 1891
Db 963 CAGCAGAGACCAAGTGTGCTTCAACACAGGATGAGGTATCATACCGCTGTGGAG 1022
QY 1892 ProSerLysProThrValLeuArgSerThrSerSerProValArgProAlaAla 1911
Db 1023 CCCAGACCGCCACGCTGCGAGGTCCACCTGCACCGTGTGGCCAGGCGCC 1082
QY 1912 ThrPheProProAlaThrHisCysProLeuGlyIleThrLeuAspGlyValTrrProThr 1931
Db 1083 ACATTTCCCACTGCGCACCACTGCGGAGGAGACCTTCATATGGGTCTTACCTTACC 1142
QY 1932 LeuMetGlnProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro 1951
Db 1143 CTATGAGCGCTCTTGTGCTCCCAAGAGGCCCGCGGCTGCGCGGCGAGCGGCC 1202
QY 1952 ArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro 1971
Db 1203 CGAGCAGACACCGGCGCATGCTTCTGCGCAAGCCCGGCGGCTGAGAGCC 1262
QY 1972 AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHis 1991
Db 1263 GCTCTCTCCCAAGAGGCTCGAGGCCCGGCGCTTGTGTGCTCTGTGCGAC 1322

QY 1992 AlaThrIleAlaArgThrProAlaIuysAenLeuAlaProHisHlaSerProaspPro 2011
| | | | |
Db 1323 GCCACCATCGCCCGGACCCCTCGAAGAACCTCGACCTCACCAACCCGCGGACCCG 1382
QY 2012 ProAlaProProAlaSerIaSerAspProHisArgGluIysThrGlnSerIysProPhe 2031
| | | | |
Db 1383 CCGGGCGCCACTGCTCGGCTCGGACCCGACCCGGAAGAAAGCTCAAGTAAACCTTT 1442
QY 2032 SerIleGlnIuLeuGluLeuArgSerLeuGluIYTHiSglYSerSerTYrSerProGlu 2051
| | | | |
Db 1443 TTCATCCAGGAACCTGAACCTCCCTTCTCTGGGTACACGACGACGACGACCCCGAA 1502
QY 2052 GluIuAlaGluProAlaSerProAlaSerSerProSerLeuThrHisAspIysGluLeuPro 2071
| | | | |
Db 1503 GGGGTGACCCCGTCAAGCCCTGTGAGCTCAACCACTGACCCACGAAGGGGCTCCCC 1562
QY 2072 IYvHIsleuGluGluLeuAspIysSerHisIleuGluGluIuLeuArgProIysGlnPro 2091
| | | | |
Db 1563 AAGCACTGGAAGAGCTGCACAAAGACCACTGGAAGGGAGCTGCGGCCAAGACCA 1622
QY 2092 GluProAlaIysLeuGluIYglIuAlaAlaHisIleuProHisIleuArgProLeuProGlu 2111
| | | | |
Db 1623 GGGCCCGTGAAGCTTGCGGGGAGGCGCCGACCTCCACACTGCGCGCTGCTGAG 1682
QY 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGluIYHisGlnArg 2131
| | | | |
Db 1683 ACCCAGCCCTCGTCCAGCCGCTGCTCCAGACCGCCGAGGGGTCAAGGTCAACCA 1742
QY 2132 ValIValThrLeuAlaGlnHisIleSerGluValIleThrGluAspIYThrArgHisSHs 2151
| | | | |
Db 1743 GTGGTACCTCGCCGACGACATCACTGAGGTATACACAGGATACACCGGACCAAC 1802
QY 2152 ProGlnGlnIleuSerAlaProLeuProAlaProLeuTYrSerPheProGluIaSerCys 2171
| | | | |
Db 1803 CCACAGCACTCAAGCCACCCCTGCGCCGCCCCCTTACTCTCTCCCTGGGGGCACTGC 1862
QY 2172 ProValIleuAspIleuArgArgProProSerAspLeuTYrIleuProProAspHisGly 2191
| | | | |
Db 1863 CCGGTCTGGAACCTCGCGCCGACCACTGATGACTTACTCTCCGCGCCGAGCATGGT 1922
QY 2192 AlaProAlaArgGlySerProHisSerGluGluIYIysArgSerProGluProAsnIys 2211
| | | | |
Db 1923 GCGCCGCGCCGTGGCTCCCGCCACAGCGGAAGGGGCAAGAGTCTCCAGAGCCAAACAA 1982
QY 2212 ThrSerValIleuGluIYglIuAspGlyIleGluProValSerProProGluIYMet 2231
| | | | |
Db 1983 AGCTGGTCTTGGGTGTGTGAGGAGCGGATTTGAACCTGTGCCCAACCGAGGCGCATG 2042
QY 2232 ThrGluProGluIYHisSerArgSerAlaValTYrProLeuLeuTYrArgAspIYgluGln 2251
| | | | |
Db 2043 ACGGAGCCAGGGGCACTCCGGAAGTGTGTGATCCGCTGTACCGGGAAGGGGAACAG 2102
QY 2252 ThrGluProSerArgMetGlySerIysSerProGluIYAsnThrSerGlnProProAlaPhe 2271
| | | | |
Db 2103 ACGGAGCCAGCGAGATGGGCTCAAGTCTCCAGGCAACACAGCCAGCGCGGCTTTC 2162
QY 2272 PheSerIysLeuThrGluSerAsnSerAlaMetValIysSerIYsIYsglnIuIleAsn 2291
| | | | |
Db 2163 TTCAGCAAGCTGACCGAGACCACTCCGCAATGTCAAGTCAAGAACGAAGATCAAC 2222
QY 2292 IYvHIsleuAsnThrHisAsnArgAsnGluProGluTYrAsnIleSerGlnProGluTYr 2311
| | | | |
Db 2223 AAGAAGCTGAACCCACAAACCGGAATGAGCTGAATATCAATATCAAGCCAGCTGGAGCG 2282
QY 2312 GluIlePheAsnMetProAlaIleThrGluTYrThrGluIleuMetPheTYrArgSerGlnAla 2331
| | | | |
Db 2283 GAGATCTTCAATATGCCCCGCAATCACCGGAACAGGCTTATGACTATAGAACCAAGCG 2342
QY 2332 ValGlnGluHisIaSerThrAsnMetGluLeuGluAlaIleIleArgIYsAlaLeuMet 2351
| | | | |
Db 2343 GTGCAAGAACATGCCAGCAACCAATGAGGCTGGAAGCCATATTTAGAAAGCACTCATG 2402
QY 2352 GlyIYvTYrAspGlnTYrGluIuSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371

Db 2402 ----- 2402
QY 2372 AsnAlaSerAlaSerIleuProAlaAlaMetProIleThrAlaAlaAspGluYArgSerAsp 2391
| | | | |
Db 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGluIYglIYglIYglIYAlaIYsValSerGluYArgProSerSer 2411
| | | | |
Db 2403 -----GTTGGCGCGGGAAGGCGCAAGGTCTCTGGCGAGACCCAGCACG 2444
QY 2412 ArgIYsAlaIYsSerProAlaProGluIleuAlaSerGluYAspArgProProSerValSer 2431
| | | | |
Db 2445 CGAAAAAGCCAAAGTCCCCGGGCCCTGGGATCTGGGGAACGGGCACCTTGTCTTC 2504
QY 2432 SerIHisSerGluGluYAspCysAsnArgArgThrProLeuThrAsnArgValTYrGlu 2451
| | | | |
Db 2505 TCAGTGCACCTCGGAGGAGACCTGCACACCGCGGACGCGCTCAACACCGCGTGGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTYrAsnProLeuIleMetArgLeu 2471
| | | | |
Db 2565 GACAGGCCCTCGTCCGCAAGGTTCACAGGCATTCCCTTCAACCCCTGATTCATGGCGCTG 2624
QY 2472 GlnAlaGluValMetAlaSerProProProProGluYleuProAlaGlySerGluYProLeu 2491
| | | | |
Db 2625 CAGGCGGGGTATGAGCTTCCACACCCCAACCGGAGCTTCCCGCGGACGGGGCCCTTC 2684
QY 2492 AlaGluProHisHisAlaTYrAspGluGluProIYsProLeuLeuCysSerGlnTYrGlu 2511
| | | | |
Db 2685 GCTGGCGGCCCAACAGCGCTGGGAGCGAGGCCAAGCCACTGCTGTGCTCGCAGTACGAG 2744
QY 2512 ThrIleuSerAspSerGlu 2517
| | | | |
Db 2745 ACACCTCCGACAGCGAG 2762

Search completed: November 12, 2005, 05:30:12
Job time : 6402 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: November 11, 2005, 09:36:02 ; Search time 1201 Seconds
(without alignments)
3429.233 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAGTWATREPRYP.....WDEPRKLLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=x1p
-Q=/cgn2_1/USFTO.spool_p/US09522753/runat_02112005_173641_20819/app_query.fasta_1.2695
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NCRME=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09522753 @CNC 1.1.186 @runat_02112005_173641_20819 -NCPRI=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -IONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents NA:*
- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/prodata/1/ina/6C_COMB.seq:*
 - 6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	12922	97.8	9053	4	US-09-976-594-306
2	4165.5	31.5	7912	4	US-09-949-016-3804
3	726.5	5.5	33529	2	US-09-144-085-3
4	718	5.4	1922	2	US-08-372-652-9
5	718	5.4	1922	5	PCT-US95-16311-9
6	698	5.3	14462	4	US-09-902-540-1090
7	698	5.3	14462	4	US-09-902-540-1090
8	689	5.2	4411529	3	US-09-103-840A-1
9	685	5.2	4403765	3	US-09-103-840A-2
10	670	5.1	44377	2	US-08-804-227C-7
11	670	5.1	44377	2	US-08-804-198-1
12	668.5	5.1	4403765	3	US-09-103-840A-2

13	665	5.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
C 14	661.5	5.0	43280	2	US-08-804-227C-1	Sequence 1, Appl1
C 15	652	4.9	13987	2	US-08-804-227C-13	Sequence 13, Appl1
C 16	648.5	4.9	50937	3	US-09-428-517-1	Sequence 1, Appl1
17	641	4.9	28958	1	US-08-258-261B-6	Sequence 6, Appl1
18	641	4.9	28958	1	US-08-456-837-6	Sequence 6, Appl1
19	641	4.9	28958	1	US-08-457-342-6	Sequence 6, Appl1
20	641	4.9	28958	1	US-08-457-646A-6	Sequence 6, Appl1
21	641	4.9	28958	1	US-08-458-076A-6	Sequence 6, Appl1
22	641	4.9	28958	1	US-08-764-233A-4	Sequence 4, Appl1
23	641	4.9	28958	1	US-08-457-335A-6	Sequence 6, Appl1
24	641	4.9	28958	1	US-08-729-214-6	Sequence 6, Appl1
25	641	4.9	28958	3	US-09-028-934-6	Sequence 6, Appl1
26	641	4.9	49377	1	US-08-764-233A-1	Sequence 1, Appl1
C 27	631.5	4.8	80161	3	US-09-036-987A-1	Sequence 1, Appl1
C 28	631.5	4.8	80161	3	US-09-370-200-1	Sequence 1, Appl1
C 29	631.5	4.8	80161	4	US-09-603-207-1	Sequence 1, Appl1
C 30	624.5	4.7	68750	3	US-09-335-409-1	Sequence 1, Appl1
C 31	624.5	4.7	68750	3	US-09-568-102-1	Sequence 1, Appl1
C 32	624.5	4.7	68750	3	US-09-567-969-1	Sequence 1, Appl1
C 33	624.5	4.7	68750	3	US-09-568-480-1	Sequence 1, Appl1
C 34	624.5	4.7	68750	3	US-09-568-486-1	Sequence 1, Appl1
C 35	624.5	4.7	68750	3	US-09-568-472-1	Sequence 1, Appl1
C 36	624.5	4.7	68750	3	US-09-567-899-1	Sequence 1, Appl1
C 37	622.5	4.7	24459	4	US-09-902-540-5004	Sequence 5004, Ap
C 38	622.5	4.7	32241	4	US-09-902-540-1247	Sequence 1247, Ap
C 39	622.5	4.7	71989	3	US-09-443-501A-2	Sequence 2, Appl1
C 40	619	4.7	47981	4	US-09-679-279-1	Sequence 1, Appl1
C 41	616.5	4.7	8438	1	US-07-945-283-1	Sequence 1, Appl1
42	614.5	4.7	15872	3	US-09-105-537-1	Sequence 1, Appl1
43	614.5	4.7	15872	4	US-09-091-609-1	Sequence 1, Appl1
44	614.5	4.7	15872	4	US-09-091-609-3	Sequence 3, Appl1
45	609.5	4.6	17622	4	US-09-902-540-1125	Sequence 1125, Ap

ALIGNMENTS

RESULT 1
US-09-976-594-306
Sequence 306, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976, 594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240, 409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 306
LENGTH: 9053
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6673549 898877.6
NAME/KEY: unsure
LOCATION: 2006, 2012
OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-306
Alignment Scores:
Pred. No.: 0
Score: 12922.00
Percent Similarity: 98.26%
Best Local Similarity: 98.18%
Query Match: 97.78%
Length: 9053
Matches: 2487
Conservative: 2
Mismatches: 10
Indels: 35
Gaps: 5
DB: US-09-522-753-5 (1-2517) x US-09-976-594-306 (1-9053)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
DB 520 ATGTGGGATTCACACAGCGCTGTGGACAGATGGAGGCGCATGAGCCCGCTACCGG 579
QY 21 ProHlaSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 580 CCCCAAGGCTTCTTCCATCCAGTGCAGATCGCCGAGCACACGAGACGTGGGCTCTCG 639
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 640 GAGTACACAGACACCTCCCGCACTATGCTCCCACTGTGCGCCGAGCTCATCTCCAG 699
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 700 CCCAGCGCGGAGGCGCTCCCTGTGTGTGAGTTCCAGCCGCGGAATGAACGGTCCAG 759
QY 81 GluLeuHisLeuLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
DB 760 GAGCTCCACCTGGCGGCAGAGTCCCACTATACCTGCGAGCTGGGGAAGTCCAGAGATG 819
QY 101 GluPheIleGlySerIysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 820 GAGTTCATTGAAAGCAAGCGCCCTCGGCTTAGAGCTGCTGCTGACCCCTGCTGCCACCG 879
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIysAspArgSer 140
DB 880 TCACCCCTGTGGCCACGGGCGCGCTGCGGGATCTGAAGACTTCACCAAGACCGTAGC 939
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrTrpAspProGluLeu 160
DB 940 CTGACCGGGCAAGCTGGACCGGTGTCTCCCGCCAGCCCGCCGACACTGAGCTGAGCTG 999
QY 161 GluLeuValProProArgLeuSerIysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 1000 GAGCTGTGTCCCGCCACGCGCTGTCCAAAGAGAGACTATCCAGAACATGGACCGCGTGAC 1059
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysValLysGlnGlnIleLeu 200
DB 1060 CCAGAGATCACCATGGTAGAGCAGACAGATCTTAACCTGAAGAAAGACGACACAGCTG 1119
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 1120 GAGGAGAGAGCTGCCAAGCGCGCCGAGCTGAAGAACCCGCTGCACCGCGCCCATCGAG 1179
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysIysAlaGluAla 240
DB 1180 TCGAAGCACCGCAGCTGCTGCAATCATCTACGACGAAACCGGAAGAGCTGAAGCT 1239
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 1240 GCACATCGGATCTGGAAAGCGCTGGGGCCCAAGTGAAGTGCCTGTGAACAACGACCC 1299
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 1300 TCCGACACCCCGGAGATCATATGAAATCAATCAACCAAGCGATGCGGAAGAAAGCTA 1359
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
DB 1360 ATCTTCTACTTCAAGAGAGAGATCACGCTCGSAAACATGGAGACGAAAGTTCTCCAG 1419
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
DB 1420 CCTATGACCAAGCTCATGAGGCGCTGGAGAAAGAAAGTGAAGCGCATTCGAAACAACCC 1479
QY 321 ArgArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 1480 CGCGCGCGGAGCCCAAGAGAGCAAGGTGCGGAGTACTTACGAAAGAGATTCCTCGAGATC 1539
QY 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgIysSerGly 360
DB 1540 CGCAAGCAGCGGAGCTGCAGAGGCCATGCAAGACAGGGTGGGCCAGCGGGGCAATGGG 1599

QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1600 CTGTCCATGTCCGCGCGCCGCGCAGCAGCAGGGGTGTAGAGATCATCATGAGCCCTCTCA 1659
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1660 GAGCAGAGAACCGGAAGAGAGATGCCACAGCTGGCCGTATGCCGCCCATGCTGTAC 1719
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1720 GACGCTGACCAAGGAGCATCAAGTTTCATCACTGAATGAAGCGGCTTATGGCCGACCCCATG 1779
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
DB 1780 AAGGTGTCAAAAGACCGCCAGGTCAATGAACATGTGAAGTGAACAGAGAGAGAGACCTTTC 1839
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1840 CGAGAGAGTTCAATGACGATCCCAAGAACTTTGGCTGATGTCATCTCTCGAGAGAG 1899
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1900 AAGACAGTGTGAGTGGGCTGCTTATTACTACTGACTGAAGAAATGAGAACTATAAG 1959
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGln---GlnGlnGlnGln 499
DB 1960 AACCTGTGTAGACGAGACTATGCGCGCGCGCAAGGCGACAGCAGACNACAGCAGCAG 2019
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGlu 519
DB 2020 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2079
QY 520 LysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValGlu 539
DB 2080 AAAAGTGAAGAGAGAGAAAGAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
QY 540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
DB 2140 AACGACAAAGAGAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2199
QY 560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
DB 2200 GAGAAAGAGGCTGTGGCTCCAAAGCCGCAAAACCTGCCAACAGCAGGAAAGACCCAAA 2259
QY 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
DB 2260 GCGCCGATCACCCGCTCAATGGCTAATGAGGCCAACAGAGAGAGGCGCATCACCCCCAG 2319
QY 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
DB 2320 CAGAGCGCGAGCTGGCTCCATGAGCTGAATGAGAGTTCTCGCTGCAAGAAAGAAAG 2379
QY 620 MetGluThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArg 639
DB 2380 ATGGAAACAGCCAAAGAAAGCTCTCTTGAAACACGCGCCGAACTGGTGGCCCATCGCCCG 2439
QY 640 MetValGlySerLysThrValSerGlnCysLysAsnMetTyrPheAsnTyrLysLysArg 659
DB 2440 ATGTGTGGGCTCCAAAGACTGTGTGCAAGTGTAAAGAACTTCTCACTCAACAAAGAGAG 2499
QY 660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAla 679
DB 2500 CAGAACTTCGATGAGATCTTTGACAGCAGCAAGCTGTGAATGAGAGAGAGAGAGAGAGAG 2559
QY 680 ArgArgLysLysLysLysAlaProAlaAlaAla-SerGluGluAlaAlaPheProProVal 699
DB 2560 CGAG 2619
QY 699 IValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluLysMetValGly 719
DB 2620 GGTGAGAGATGAGAGATGAGAGGCGTCCGGCGTGAAGGAGAAATGAGAGAGAGATGTTGGA 2679
QY 719 uGluValGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyTrp 739

Db 2680 GAGAGCTGAA----- 2689
Qy 739 oAlaThrValAspAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaI 759
Db 2690 -GCCACTGTCAACAACAGCTCAGACCCGAGAGCATCCCTCTCTCTCACTGAGGCGC 2748
Qy 759 alyAspThrGlyGlnAsnGlyProlybProProAlaThrLeuGlyAlaAsp-GlyProP 779
Db 2749 CAAGGACACAGAGGACGATGGGCCCAAGCCCCCAGCCACTCTGGCGCCGACGGGGCCAC 2808
Qy 779 rOProGlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrPro 798
Db 2809 CCCGAGGCGCACCCACCCACACCGAGGACATCCGCGCC-CCCACTGAGCCACCGCG 2867
Qy 799 AlAserGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaPro 818
Db 2868 GCCTCTGAGAGCCAGGAGCCCTAGCCGCCACACAGACCCCATGCTGTCACCT 2927
Qy 819 ProProValValProlybGluGluGlyGluGluGluThrAlaAlaProProValGlu 838
Db 2928 CTTCTGTGTGTCCCAAG 2987
Qy 839 GluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 858
Db 2988 GAGGGGGAG 3047
Qy 859 GluGluProValIlybSerGluGlyThrGluGluGluGluGluGluGluGluGluGlu 878
Db 3048 GAGAGAGCCCTCAAGAGCGAGTGACCGAGAGAGCCAGAGAGGGCGCGCCAAAGGGCAAG 3107
Qy 879 AspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuAlaGluGlybGluGluGly 898
Db 3108 GAGCGAGAGCCCTGAGGCCACCGCCGAGGGCGCTCAAGAGAGAGAGAGAGAGAGAG 3167
Qy 899 GlySerGlyAlaArgAlaThrThrAlaIlybSerSerGlyAlaProGluAspSerAspSer 918
Db 3168 GGGAGGCGGAGGCGCACACAGCCAAAGAGCTCGGGCGCCCGCCGAGAGAGAGAGAG 3227
Qy 919 AlaThrCysSerAlaAspGluValaAspGluAlaGluGluGlyIlybAspAlaGluLeu 938
Db 3228 GCTACCTGAGTGACAGCGAGTGATGAGAGCGCAGGGCGGCGCAAGAACCGGCTGCTG 3287
Qy 939 SerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
Db 3288 TCCCAAGGCGCCAGCCTCTCACTCGGAGTGGGAGCCCGCGGCAATGCTCACTCCAG 3347
Qy 959 LysProLeuAspLeuIlybGluGluGluGluGluGluGluGluGluGluGluGluGlu 978
Db 3348 AAGCCACTGACCTGAAGCAGCTGAAGCAGCAGCGGCTCCATCCCCCATC---GTC 3404
Qy 979 ThrIlybValHisGluProProArgGluAspAlaAlaProThrIlybProAlaProPro- 998
Db 3405 ACCAAAGTCCATAGCCCCCGGAGAGAGCGAGCTCCCAAGAGCGAGCTCCCGCAGAC 3464
Qy 998 aProProProProGluAsnLeuGlnProGluSerAspAlaProGluGlnProGlySerSe 1018
Db 3465 CCAACCGCCACCCCAAACTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524
Qy 1018 rProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3525 CCCCCGGGGAG 3584
Qy 1034 -----AlaPhaAlaAlaGluAlaGluGluGluGluGluGluGluGluGluGluGlu 1050
Db 3585 GTTCTTCCCAAGCCTTGACAGCGAGAGCGCAAGAGCTGCTGGGAGACCCCTTGCTGAGAC 3644
Qy 1050 rSerGlyLeuProPhaProValProProArgGluValIleLysAlaSerProHisAlaPr 1070
Db 3645 TTCGGGCTGCTCTTCCCGGTGCCCCCGTGAGGTGATCAAGGCGCTCCCGCATGCCCC 3704
Qy 1070 oAaPProSerAlaPhaSerThrAlaProProGlyHisIlybProLeuProLeuGlyLeuHis 1090

Db 3705 GAGCCCCAGCCTTCTCTACGCTCCAGCTGTACCCCACTGCGGCGCTCATGA 3764
Qy 1090 pThrAlaArgProValLeuProArgProProThrIleSerAspProProLeuIleSe 1110
Db 3765 CACTGCCCCGCGCTCTGCGGCGCCACCCACCATCTCCAAACGCGCTCCCTCATCTC 3824
Qy 1110 rSerAlaLysHisIlybProSerValLeuGluArgGlnIleGlyAlaIleSerGlnIlyMeSe 1130
Db 3825 CTCTGCCAAGACACCCAGCGCTCTCGAGAGGCAAAATGATGCCATCTCCAAAGGAATGC 3884
Qy 1130 rValGluLeuHisValProTybSerGluHisAlaLysAlaProValGlyProValThrMe 1150
Db 3885 GGTTCACCTCCAGCTCCCGTATCTCAGAGCATGCCCCAAGGCCCGGTGGGCGCTGTCA 3944
Qy 1150 rGlyLeuProLeuProMetAspProlybLysLeuAlaProPhaSerGlyValLysGlnI 1170
Db 3945 GGGGCTGCCCCCTGCCATGAGACCCCAAAAGCTGGCACCTTCAAGCGAGAGAGAGAG 4004
Qy 1170 uGluLeuSerProArgGlyGluAlaGlyProProGluSerLeuGlyValProThrAlaGlu 1190
Db 4005 GAGCTGTCCCAAGGGGCCAGGCTGGCCACCGAGAGAGCTGGGGGTGCCACAGCCCA 4064
Qy 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db 4065 GAGAGCGTCCGTCTGAGAGGAGACAGCTTGAGCTCAGTTCCGGGCGGAAGATCACAA 4124
Qy 1210 rGlyIleProSerThrArgValProSerAspSerAlaIleThrTybArgGlySerIleTh 1230
Db 4125 AGGCATTTCCAGACACAGGGTGCCTTGACAGGCGCATCAATACCGGCTCATCAC 4184
Qy 1230 rHisGlyThrProAlaAspValLeuTybLysGlyThrIleThrArgIleIleGlyLys 1250
Db 4185 CCAAGGAGCCCAAGCTGACCTCTCTCAAGAGGACCATACCAAGATCATCGCGAGGA 4244
Qy 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProGlyHisValIleTy 1270
Db 4245 CAAGCGAGTGGCTTGAGACGGCGCGGAGAGACAGCTGCCCAAGGGCACGTCATCTA 4304
Qy 1270 rGluGlyLysLeuGlyHisValLeuSerTybGluGlyGlyIlybMetSerValThrGlnCybSe 1290
Db 4305 CGAAGCGAAGAGGCGCACCTTGTCTATGAGGGTGATGTGTGACCCAGTGCTC 4364
Qy 1290 rLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 4365 CAAGAGAGAGCGAG 4424
Qy 1310 rTybAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 4425 CTATGACATGATGAGAGGCGCGCTGGGAGAGAGCATCTCTCAGCCAGCATCGAAGGTC 4484
Qy 1330 uMetGlyArgGluAlaProProGluArgHisSerProHisIleLeuGluGlnHisHis 1350
Db 4485 CATGGGCGGCGCATCCCGCGAGGAGACACAGCCCAACCATCAAAAGAGAGAGAGAG 4544
Qy 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTybValGluAlaGlnGluAspTy 1370
Db 4545 CATCCCGGGTTCATACACAGAGGATCTCTCGGTCTCTAGTGAAGGACAGAGAGACTA 4604
Qy 1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSe 1390
Db 4605 CTTGCTCGGAGGCGCAAGCTCTTAAGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4664
Qy 1390 rArgAspLeuThrGluAlaTybTybThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 4665 ACGGAGCTGACCGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4724
Qy 1410 rHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 4725 CCATGAGGGCGCTGTGCGCACCGGTGAAGAGGGCGGCGCTCATCATGAGATCCCGG 4784
Qy 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySe 1450
Db 4785 CGAGAGCTGCGGACACCGCCGAGAGCTGCCCTTGCGCCGCGGCGCTCAAGAGAGGCTC 4844

QY 1450 rllertHrGInGlyThrProLeuLysTyrAspThrGlyValAspThrThrGlySerLysTyr 1470
DB 4845 CATCACGACGAGGACCCCTCATAGTACACACCGCGCTCCACCATCTGCCTCCAAAAA 4904
QY 1470 gHIsaAPValArgSerLeuIleGlySerProGlyValArgThrPheProProValHisProLe 1490
DB 4905 GCACGACGTACCTCTCCATCGGCAGCCCGCGCGAGCGTTCCACCCGTGCACCCGCT 4964
QY 1490 uAspValMetAlaAspAlaArgAlaLeuGluuArgAlaCysTyrGluGluSerLeuLysSe 1510
DB 4965 GGAATGTAATGGCCGACGCGGCGACTGGAACTGCTGTCTATGAGAGAGAGCTGAAGAG 5024
QY 1510 rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyValaProValIleVal 1530
DB 5025 CCGGCGACAGGACCGCGACGAGCTCGGGGGCTCATTTGGCGCGGCGCCCGTATTGT 5084
QY 1530 lProGluLeuGlyLysProArgGluInserProLeuThrTyrGluAspHisGlyAlaProPh 1550
DB 5085 GCCTGAGCTGGGACAGCCCGCGACGAGCCCTTAACCTATGAGGACACCGGGGCAACCTT 5144
QY 1550 eAlaGlyHisLeuProArgGlySerProValThiMetArgGluProThrProArgLeuG 1570
DB 5145 TGCCTGGCCACTCCACGAGAGTTGCCCTGTACACGCGGAGCCCAAGCCGCGCTTGA 5204
QY 1570 nGluGlySerLeuSerSerSerLysAlaSerGluAspArgLysLeuThrSerThrProAr 1590
DB 5205 GGAAGGCAACCTTTCCTCCAGCAAGGCATCCACGACCGAAAGCTGACCGTCCAGCTCG 5264
QY 1590 gGluIleAlaLysSerProHisSerThrValaProGluHisHisProHisProIleSerPr 1610
DB 5265 TGGAGATCGGCAAGTCCCCGACAGCACCGTGCCTCGGACACCCACCCCTATCGGCC 5324
QY 1610 oTyrGlyHisLeuLeuArgGlyValaSerGlyValaAspLeuTyrArgSerHisIleProle 1630
DB 5325 CTATGAGCACTGCTTCGGGCGTAGAGTGGCGTGAACCTGTATCGAGGCCACATTCCTC 5384
QY 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIaTyrTy 1650
DB 5385 GGCCTTCGAGCCCACTTCATACCCCGGCGCATCCCTGTGAGCGACGCGCTCTACATA 5444
QY 1650 rLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuI 1670
DB 5445 CCGGCCCAACACTGGCCCCCAACCCCACTACCCGCACTGTACCCACCTCACTCAT 5504
QY 1670 eArgGlyTyrProAspThrAlaAlaLeuGluuAsnArgInThrIleIleAsnAspTyrI 1690
DB 5505 CCGCGGCTACCCCGACACGCGCGGCTGGAGAACCGGACAGACCATCATATGACTACAT 5564
QY 1690 eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgHisAspMetLe 1710
DB 5565 CACCTCGACGACGATGCACACAAACGGGCGCACCGGCATGGCCCAAGGAGCTATATGCT 5624
QY 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAspTyrAlaAlaGlyProArgG 1730
DB 5625 GAGGGGCTCTCTCGCCCGGAGTCTCTGTGGCACTCACTACCTGCGGGTCCCGAGAG 5684
QY 1730 yIleIleAspLeuSerGlnValaProHisLeuProValIleuValaProProThrProGlyTh 1750
DB 5685 CATCATGACCTGTCCCAAGTGCACACCTGCTGTGCTGTGCCCCCGACACAGGACAC 5744
QY 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaAProGlnProPheSerSe 1770
DB 5745 CCAAGCACCGGCATGAGACCGCTTGCTACCTCCCAACCGGCGCCCAAGCCTTCAGCAG 5804
QY 1770 rArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
DB 5805 CCCCCACAGAGCTCCCACTCTCTCCAGAGAGGTCCAAACACACTTGCAAAACCAACAC 5864
QY 1790 rThrArgSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArg 1810
DB 5865 CACGCTCTGCTCGAGCGGAGCGAGACCGGAGTCGAGAGCGGAGACCGGGATCGGAGCG 5924

QY 1810 gGluLysSerIleLeuThrSerThrThrValaGluHisAlaProIleTPrArgProGl 1830
DB 5925 GGAATAAGTTCATCTCACGTCCACACGCGGTGAGCAGCAGACCATCTGTGAGACTGG 5984
QY 1830 yThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerAr 1850
DB 5985 TACAGAGCAGACACGGGAGGAGCGGACGAGCGGCGGGGGTGGGGGAGACAGCAGCGG 6044
QY 1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
DB 6045 CCCCCCTCCCATCTCCATGCCACAGCATCTGCCCATCTCCCTCGGACCCAGATGC 6104
QY 1870 aLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVa 1890
DB 6105 CCTCAGCAGAGACCCAGTGTCTTCACAAACACAGCAAGAGGGATGATATACCGCTGT 6164
QY 1890 lGluProSerLysProThrValLeu-----ArgSerThrSerThrSerSerProVa 1907
DB 6165 GGAAGCCAGACAGCCCAAGTCTTGAAGTGGGCGAGGTCCACTTCACCTCTCCACCGCT 6224
QY 1907 lArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGl 1927
DB 6225 TCGCCGGCTGCACATTCACACTGCACCCACTGCGCCACTGGGGGACCCCTGATGC 6284
QY 1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValaAlaAr 1947
DB 6285 GGTCTTACCTTACCTCATGAGACCGGTCTTGCTGCTCCAGAGAGGCCCCCGGGGTGCCG 6344
QY 1947 gProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
DB 6345 GCGAGAGCGGCCCCGAGCAGACCGGCGCATCTCTCTCGCCAGCCCCCAGCCGCTC 6404
QY 1967 rGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValaProPr 1987
DB 6405 CCGGCTGAGCCCGCTCTCCCGCCAGCAAGGAGCTCGAGCCCCGCGCTGAGCTCC 6464
QY 1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAl 2007
DB 6465 TGTCTGTGGCCAGCGCACCATGCGCGCACCCCTGGAGAAACCTCGACCTCACACCC 6524
QY 2007 aSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGl 2027
DB 6525 CAGCCCGAGCCCGCGGCCCACTGCTCGGCTCGGACCCGACCGGAGAAAGACTCA 6584
QY 2027 nSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
DB 6585 AGTAAACCTTTTCATCATCAGAACTGGAACTCCGTTCTGTGGTTACACAGCGACGAG 6644
QY 2047 rTyrSerProGluGlyValaGluProValaSerProValaSerSerProSerLeuThrHisAs 2067
DB 6645 CTACAGCCCCGAAAGGGGTGAGACCCGTCAAGCCCTGTGAGACTCACCGATCTGACCA 6704
QY 2067 pLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGlyLeuArg 2087
DB 6705 CAAGGGGCTCCCAAGCACCTTGAGAGCTTCAGAAAGACCACTGAGAGGGAGCTGGG 6764
QY 2087 gProLysGlnProGlyProValaLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr 2107
DB 6765 GCCCAAGCAGCAGGCGCCGCTGAGAGCTTGGCGGAGGCGCCCACTCCCACTGCGG 6824
QY 2107 gProLeuProGluSerGlnProSerSerSerProLeuLeuGlnInThrAlaProGlyVala 2127
DB 6825 GCGGCTGCTGAGAGCAGACCCCTGCTCAGCCCGCTGTCCAGACCGGCCCAAGGGGTCA 6884
QY 2127 gGlyHisGlnArgValaValaThrLeuAlaGlnHisIleSerGluValaIleThrGlnAspTy 2147
DB 6885 AGGTACACAGCGGGTGTACCTTGCCCAAGACATCAAGTAGTATACACAGAGACTA 6944
QY 2147 rThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePr 2167
DB 6945 CACCCGACACACCCACAGAGCTGAGGACCCCTGCGGCGCCCTCTACTGCTTCCC 7004
QY 2167 oGlyAlaSerCysProValaLeuAspLeuArgArgProProSerAspLeuTyrLeuProPr 2187

DB 7005 TGGGGGCGAGCTGCCCCCTCTCCGAGCTCCGCGCCGACCAAGTACCTTACCTCCGCGC 7064
QY 2187 OPRoAAPHIaGlyValaProAlaArgLySerProHiseSerGlyValaLysaArgSerPr 2207
DB 7065 CCGGACCAATAGTGGCCCCGCGCTGGCTCCCGCCACAGCAAGAGGGGCAAGAGTCTCC 7124
QY 2207 OGluProAsnLysEthSerValLeuGlyGlyGlyValaAspGlyLileGluProValaSerPr 2227
DB 7125 AGAGCCAAACAAGACGTGCTTGGGTGGTGGTGAAGAGCGATTGAACCTGTGTCCC 7184
QY 2227 OPRoGluGlyMetThrGluProGlyHiseSerArgSerAlaValaThrProLeuLeuYrAr 2247
DB 7185 ACCGGAGGCGATACGAGGACGAGGCACTCCGAGAGTGTGTGAACCGCTGTACCG 7244
QY 2247 GAAPGlyGluGluThrGluProSerArgMetGlySerLysSerProGlyLeuThrSerG 2267
DB 7245 GGAATGGGGAACAAGACGAGCCGACAGATGGCTCCAAAGTCTCCAGCAACACAGCCA 7304
QY 2267 nPRoProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLy 2287
DB 7305 GCGGCGAGCTTCTTCCAGCAAGCTGACCGAGAGCACTCCGCGCATGTCAAGTCCAGAA 7364
QY 2287 BGInGluLileAsnLysLysLeuLeuThrHiseAsnArgAsnGluProGluThrAsnLise 2307
DB 7365 GCAAGAGATCAACAAGAACGCTGAACCCACCAACCGGAATGAGCTGAATACATATCAG 7424
QY 2307 rGInPRoGlyThrGluLilePheAsnMetProAlaLileThrGlyThrGlyLeuMetThrTy 2327
DB 7425 CCAAGCTGGAGCAAGATCTTCAATATGCCCCGCTCAACGGAACAGGCTTATGACCTA 7484
QY 2327 rASerSerGlnAlaValaGlnGluHiseLysaSerThrAsnMetGlyLeuGluLalaLileLea 2347
DB 7485 TAAaACCGAGCCGCTGCGAGAACATGCGACACCAACATGGGCTGAGGCGCTAATTAG 7544
QY 2347 GlyValaLeuMetGlyLysaThrAspGlnThrGluGluSerProProLeuSerAlaAsnAl 2367
DB 7545 AAAGGACATCATGGTAATATATGACAGTGGAAAGATCCCGCGCTCAGCGCAATGC 7604
QY 2367 aPHeaPRoLeuAsnAlaSerLysaSerLeuProAlaAlaMetProLileThrAlaLaa 2387
DB 7605 TTTTAAACCTCTAATATGCGAGTCCAGCTGCGCGCTGTATGCCATTAACCTCTGTGA 7664
QY 2387 rGlyASerSerAspHiseThrLeuThrSerProGlyGlyGlyValaLysaValaSerG 2407
DB 7665 CGGACCGAGTGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7724
QY 2407 yArGPRoSerSerArgLysaLysaSerProAlaProGlyLeuAlaSerGlyAspArgPr 2427
DB 7725 CAGACCCAGAGCCGAAAGCCAAAGTCCCGCGCGCTGGCATCTGGGAGACCGGCC 7784
QY 2427 oPRoSerValaSerSerValaHiseSerGluGlyAspArgAsnArgThrProLeuThrAs 2447
DB 7785 ACCCTGTCTCTCAAGTGCATCGAGGAGACTGCAACCGCGGACCGCTCCACAA 7844
QY 2447 nArGValaThrGluAspArgProSerSerAlaGlySerThrProPheProTyranProle 2467
DB 7845 CCGCGCTGGAGAGACAGCCCTCTGCTCCGAGTTTCCAGCCATTCCCTTACAACCCCT 7904
QY 2467 uLleMetArgLeuGlnAlaGlyValaMetAlaSerProProProProGlyLeuProAlaG 2487
DB 7905 GATCATGCGGCTGCAAGCGGGGTGTCAATGGCTTCCCAACCCGAGGCTCCCGGG 7964
QY 2487 ySerGlyProLeuAlaGlyProHiseLysaLysaArgGluGluProLysProLeuLeuY 2507
DB 7965 CAGCGGCGCTCTGCTGGCGCCCAACAAGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 8024
QY 2507 sSerGlnTyGluThrLeuSerAspSerGlu 2517
DB 8025 CTCGCACTAGACACTCTCCGACAGCGAG 8055

/ Sequence 3804, Application US/09949016
/ Patent No. 6812319
/ GENERAL INFORMATION:
/ APPLICANT: VENTNER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949, 016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ. ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ. ID NO 3804
/ LENGTH: 7912
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-3804

Alignment Scores:

Pred. No.:	9,68e-183	Length:	7912
Score:	4165.50	Matches:	1103
Percent Similarity:	53.57%	Conservative:	344
Best Local Similarity:	40.84%	Mismatches:	777
Query Match:	31.52%	Indels:	477
DB:	4	Gaps:	109

US-09-522-753-5 (1-2517) x US-09-949-016-3804 (1-7912)

QY 16 GluProArgTyProProHiseSerLysaSerTyProValaGlnLileAlaGlyThrHiseThr 35
DB 289 CAAGTCTGATCTCTCTCACTGTGCAAGTATATCCCAACACCCGCGCACAGCAG 348
QY 36 AspValaGlyLeuLeuGluGluTyGln-----HiseLysaArgAspTyraLaseRHisLeu 53
DB 349 GAGTTGGCAGCTCCCTGATATTCGTTCTCTCATCTTGAAGTGAAGTCAAGCATCAGCTT 408
QY 54 SerProGlySerLileGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73
DB 409 TTGCAGCAACAGCAGAGCAAGCTTCGAAAGCGACCTTCTTGGCTTTCAGAAATTTTCA 468
QY 74 ProGlyAsnGluArgSerGlnGluLeuHiseLeuArgProGluSerHiseSerTyrLeuPro 93
DB 469 CCAAGTCTGACAGCGCTCAAGAA-----AGGAGAACTAATTATGAAACCGTTTCATCCA 522
QY 94 GluLeuGlyLysaSerGluMetGluPheLileGlySerLysaArgProArgLeuGluLeu 113
DB 523 GCGCCATCCCGACGATCATGATTCATGAAATCAAGAGCAAGCTCTGGAAACAGTT 582
QY 114 ProAsp-----ProLeuLeuArgProSerPro 122
DB 583 TCGATTCATCATTTTCAGCGGTGCAAGTGGCTTTGCTTTAAGCAACCGCTGCCA 642
QY 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysaAspArgSerLeuThr 142
DB 643 -----GAAGGCTGAGGCT---TCTGCAGATGCTAAGAAAGATCCAGCATTCGGA 690
QY 143 GlyLysaLeuGlu---ProValaSerProProSerProProHiseThrAspProGluLeuGlu 161
DB 691 GCGAAACATGAGACTCATCTCTCCAAATTTGGGGAACATGTCGAGATGATCAAAAT 750
QY 162 LeuValaProProArgLeuSerLysaGluLeuLileGlnAsnMetAspArgValaAspArg 181
DB 751 GCTTCACTTCAAAACTTCAAGGAAGATTAATACAGATATGATGATGATGATGATGAT 810
QY 182 GluLileThrMetValaGluGlnGlnLileSerLysaLeuLysaLysaGlnGlnGlnGln 201
DB 811 GAAATTCGAAAGATGAAACAGAGATCTTAAATGAAAGAAACAAACAGCTTGA 870

QY 202 GIUGLUALAAlaIaIySerProGluProGluIySProVaISeSerProProIleGluSer 221
 DB GAAGAGGCGAGCTTAACCTCTGAGCGCTGAGAGCCCGTCTCCCTCTCTGAGGAGCG 930
 QY 222 LysH1aH3SerLeuValGlnIleIleIyTAspGluAsnArgIyValaGluIaIaIa 241
 DB 931 AAACACCGGAGATTTGTCMAAATTATTATGATGAGAAATCGAAMAAACAGAGAAAGCT 990
 QY 242 H1aA3G1LeuG1GluIyLeuG1yProG1nValaG1LeuProLeuTyTAsnG1nProSer 261
 DB 991 CATMAAATTMTTGAAGGCTTGGCCCAAAAGTTGAATCCCACTGATTAACAGCCATCA 1050
 QY 262 AspThrArgIlnTyH1sG1uAsnIleIyS1IeAsnG1nAlaMeArG1yS1eIle 281
 DB 1051 GATACCAAGGTGACATGAGAAATCATCAAGACAAACAGGTGATGAGAAAAATCATTT 1110
 QY 282 LeuTyTPhelYArgArgAsnH1aIaArgIyG1nTPuIyG1nIyS1heYSeG1nArg 301
 DB 1111 TTATTTTAAAAAGAAATCATGCAAGAAACAAAGGAAACAAAAATCTGCCAGCGT 1170
 QY 302 TyTAspG1nLeuMeG1uAlaIeugIuIyS1yValaG1uArgI1eG1uAsnAsnProArg 321
 DB 1171 TATGATCACTCATGAGGCAATGGGAAAGAAAGTGAAGAAATGAGAAATTAATCTCGG 1230
 QY 322 ArgATgAlaIeYSeIySerIySValaArgIuTyTArgIuIySg1nPhProG1uIleArg 341
 DB 1231 AGGAAGCTTAAGAAAGCAAAACAAAGGAATCTATGAAAGACAGTTTCAGAAATTCGA 1290
 QY 342 LysG1nArgIuIeugIuIyArgMeG1nIySerArgValaG1uArgIySergIyLeu 361
 DB 1291 AAACAAAGGAAACAGAAAGAAAGATTTCAG--CGAGTTGGCAGAGGGAGCTGATCTT 1347
 QY 362 SerMeTSeR1aIaIaArgSerG1uH1sG1uValaSeRg1uIleIeAspG1yLeuSerG1u 381
 DB 1348 TACGCCACCATGCTCTAGAGTGAAGCATGAGATTTCGAATTAATTAATGAGCTCTGAG 1407
 QY 382 G1uIuAsnLeuG1uIySg1nMeArG1nLeuAlaValIleProMeLeuTyTAsp 401
 DB 1408 CAGGAATTAATGAGAAACAAATGCGGAGCTCTGTGATTCACATTAATGATTTGAT 1467
 QY 402 AlaAspG1nArgI1eIyS1heIleAsnMeTAsnG1yLeuMeTAlaAspProMeIyS 421
 DB 1468 GCAGAAACAAAGCAGACTCAAGTTCATTAACATGAAGGGCTTAATGAGAGACCTTAAGA 1527
 QY 422 ValIyTyrLeuAspArgIlnValMeTAsnMeTTPSeRg1uG1nIySg1uThrPheArg 441
 DB 1528 GTGTATTAAGATAGGCAATTTATGAATGTTGACATGACCATGAAGAAAGACCTTAAG 1587
 QY 442 G1uIySPhemeG1nH1sProIySAsnPhelYLeuIleAlaSeRPhelueG1uArgIyS 461
 DB 1588 GACAACTTATCCAGCATCCAAAAAATTTGACTAATTCATCATTAATCTTGAGAGAGAA 1647
 QY 462 ThrValAlaG1uCyValLeuTyTyrTyTLeuThyIySAsnIuIyAsnTyTlySser 481
 DB 1648 AGGTTCCTGATTTGTTTGTATTACTATTAAACAAGAAATATGAATTAAGGCC 1707
 QY 482 LeuValaIArgSerTyT--ArgATgArgG1yLysSerG1nG1nG1nG1nG1n 500
 DB 1708 CTCGTAGAAAGATTAATGAGAAACCAAGGCAAGCAACCAAGAAATGCTGACACCTCG 1767
 QY 501 G1nG1nG1nG1nG1nG1nG1nG1nG1nG1nPrometProArgSerSerG1nG1nG1n 520
 DB 1768 CAGAAGAAAGTAAGAAAGAAAGAAAGAG--GATTAACAGCAAAAGAAAGCAAGAA 1821
 QY 521 AspG1uIyG1uIySg1uIySg1uIaG1uIySg1uIySProG1uValaG1uAsn 540
 DB 1822 AAAGAAAGAAAGAAAGAAAGTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGTGA 1881
 QY 541 AspIySg1uIySg1uIySg1uIySg1uIySg1uIySg1uIySg1uIySg1uIySg1u 560
 DB 1882 ACCAAGAA-----AAGCAAGATAGATGATGACCA--GAAGAAATGAGAGAA 1929
 QY 561 LysG1uAlaValaIaSerIySg1yArgIyS1ThrAlaAsnSerG1nIyArgArgIySg1y 580

DB 1930 AGAGAGCAAGCCACACCCCGGGGGCGAAAGACTGCAACAGTCAGAGCGCGGTAAAGGC 1989
 QY 581 ArgIleThrArgSerMeTAlaAsnG1uAlaAsnSerG1uG1uAlaIleThrProG1nG1n 600
 DB 1990 CGGATACCAAGATCCATGACAAACAGAGCTGACAGCTGCCAGTGCTGCAGCCAGCGGT 2049
 QY 601 Ser-----AlaG1uLeuAlaSerMeG1uLeuAsn 610
 DB 2050 ACTGAAGAGCCCCCAACCTCTGCCACCGCCACAGAAACCATTTCTACAGAGCTGTG 2109
 QY 611 GluSerSerArgTTPThG1uG1uIyMeTg1uThrAlaIyS1yG1yLeuLeuG1uH1s 630
 DB 2110 GAGACCTCTCGATGAGCAAGAAAGAAATGAAAGTTGCTAAAGAAAGGTCTATGAGAACT 2169
 QY 631 G1yArgAsnTPSeR1aIleAlaArgMeTValaG1ySerIyS1ThrValSerG1nIyS1yS 650
 DB 2170 GGTCTTAATCTGAGCAATTCCTTAATATGTTGGAAACGAAAGTGAAGCTCAATGTA 2229
 QY 651 AsnPhetyTyrPheAsnTyTlyS1ySArgIlnAsnLeuAspG1uIleIeugIlnIyS1yS 670
 DB 2230 AACTTCTATTATTAATCAATAAAGCGACACAAATCTTGACAACTCTTACAGCAGCATAA 2289
 QY 671 LeuIySMeTg1uIySg1uArgAsnAlaArgArgIyS1yS1yS1yS1yS1yS1yS1yS 690
 DB 2290 CAGAAACTTCACGAAACCTCTGTAAGAGCAGATGTCTCAATGTGAAGTGTGCT 2349
 QY 691 SerG1uG1uAlaAlaIlePheProProValaIaG1uAspG1uIySg1uIySg1uIySg1yVal 710
 DB 2350 TCCACTGTTTCTGCT-----CAGAGATGAAGATATGAGACTCC----- 2391
 QY 711 SerG1yAsnG1uG1uIySg1uMeTValaG1uIySg1uAlaIeUHIaIaSeRg1yAsnG1u 730
 DB 2392 -----AATGAAGAAAGAAATCCAGAAAGCACAGAA----- 2421
 QY 731 ValProArgG1yG1uCySerSerG1yProAlaThrValaAsnSerSerAspThrG1uSer 750
 DB 2421 ----- 2421
 QY 751 IleProSerProH1sThrG1uAlaIySAspThrG1yG1nAsnG1yProIySProPro 770
 DB 2422 -----CTGAACCTGTCAAG----- 2436
 QY 771 AlaThrLeuG1yAlaAspG1yProProProG1yProProThrProProArgThrSer 790
 DB 2437 -----CCACGAGAGCAGCTCGAAATGCTACTTCT 2469
 QY 791 ArgAlaProIleG1uProThrProAlaSerG1uAlaThrG1yAlaProThrProProPro 810
 DB 2470 CGAGGAACACAGAACCTGCGGTGAGCTTGAG-----CCACACAGGAAACT 2517
 QY 811 AlaPro--ProSerProSer--AlaProProProValaIySg1uIySg1uIySg1u 828
 DB 2518 GACCCAGTACATCTCCCTCTTACAGTTCCAAGTCAAAACCAAGCTGAAGTGAAGT 2577
 QY 829 G1uG1uThrAlaAlaIa 837
 DB 2578 GTGAGAGCCAGGTGAATGACAGCATCACTGCTGAGACAGACAGACAGATGATGAT 2637
 QY 838 -----G1uG1uG1yG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uG1u 851
 DB 2638 CAGCAGAGACACAGTCTGMAAGAGGTCTGTTGTGATCCCAACCCCTCAACCAAA-- 2694
 QY 852 AlaValAspThrG1yLySAlaG1uG1uIySg1uIySg1uIySg1uIySg1uIySg1uIySg1u 871
 DB 2695 ---GCTGACTCTGTGACCTTAAGTGAAGAGGTGCCAGAAACATCATCAATGAAGTGA 2751
 QY 872 G1uG1yProAlaIySg1yLySAspAlaG1uAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 891
 DB 2752 GGTGATTAATCAAGAAAGAGACTTGATAGAGCC----- 2787
 QY 892 LysAlaG1uIySg1uIySg1uIySg1yArgAlaThrThrAlaIySserSerG1yAla 911

Db 2788 ---AGTAGAAGGTGGAACCTTAGATGAAGATTGTTGTTAGCTCAGCAATAATTAATGCC 2844
Qy 912 -----ProGln---AapSerAapSerSerAlaThrCysSerAlaAapGluVal 926
Db 2845 CAAAGGCCCGAGGCCGACATGACAAATGATTTCCAGTGCACGTCGACGCTGATGAG--- 2901
Qy 927 AapGluAlaGluGlyGlyAapIysAaenAArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGTGAGAGCAGAGGACAGAGAATGTTCTTATGAGACTCAAGACCTTCA 2958
Qy 944 LeuLeuThrProThrGlyAapProAlaAlaAaenAlaSerProGlnIys-----ProLeu 961
Db 2959 CTGTTAAACCCCACTGATCTTACTACTGTC---TCATCTCCGTTAAACCAATTCACATG 3015
Qy 962 AapLeuLeuGlnLeuIysGlnAArgAlaAlaAlaProProIle----- 976
Db 3016 GATCTGCCACAGCTTACAGATCGAGCTGTGTTATCCACCAATGATTCCTGCACCCCA 3075
Qy 977 -----GlnValThr 979
Db 3076 TGTAAATACCAATTGGAACCCAGTAGCGGCTATGCTCTTACAGAGGACACATTAA 3135
Qy 980 IysValAlaGluProProAArgGluAapAlaAlaProThrIysProAlaProProAlaPro 999
Db 3136 GCATGTGCATGAGTACAGCACTCTCTGAG----- 3162
Qy 1000 ProProProGlnAaenLeuGlnProGlnIysAapAlaProGlnGlnProGlnIysSerPro 1019
Db 3163 ---GACACAGGAGAGACAGACAAAGATGATTTGGAATGTAAGATTCTTCAAACTTCCA 3219
Qy 1020 AArgIlyLysSerAArgSerProAlaProProAlaAapIysGluAlaAaenAlaAlaGluAla 1039
Db 3220 TGTGGACATCCAGAGGTCCA-----AACAGAGAG----- 3249
Qy 1040 GlnIysLeuProGlyAapProProCysTrpThrSerGlyLeuProProProAlaProPro 1059
Db 3250 -----TGG----- 3252
Qy 1060 AArgGluValIleLysAlaSerProAlaAlaProAapProSerAlaAapSerIlyAlaPro 1079
Db 3253 ---GAAGTCTTTCAGCTCTCCACAT-----CAAGTGAATTAATCTTCCCT 3297
Qy 1080 ProGlnIysAapLeuProLeuGlnIysAaenAlaAapThrAlaAArgProValLeuProAArgPro 1099
Db 3298 GAAGGCTTCCGGTTCCG-----ACAACTGCAGCAACAGGACCCAGCCGCC 3342
Qy 1100 ProThrIleSerAaenProProProLeuIleSerSerAlaLysIleProSerValLeuGln 1119
Db 3343 CCTCTCATCCCGTCAATCCAAACCAACAGGTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
Qy 1120 AArgGlnIleGlyAlaAlaIleSerGlnIly-----MetSerValGlnLeuHis 1134
Db 3397 ---ATGGAGGAGCTTCATCTCACAGGGAACACAGCACTTATTGACTTCTCATTAATCAG 3453
Qy 1135 ValProIysSerGlnIleAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db 3454 GCTTCTTCACTCAAGAAACACCCAGCCGTCAGTGGATCTATCTCTTGGACCTGCCA 3513
Qy 1154 LeuProMetAapProIlyLysLeuAlaProPheSerGlyValIlyGlnGlnLeuSer 1173
Db 3514 CCGCAACAGGAATCTGCCAAATCATCTTGGCTTACCTTACATCAAGCAGAGAAATTTTCT 3573
Qy 1174 ProAArgIlyGlnAlaGlyProProGlnIysLeuGlyValProThrAlaGlnGluAaSer 1193
Db 3574 CCCCAGAACCAAACTCACACCTGAGGCTGTGTGTC---AGGCCCCAACATGAAAGT 3630
Qy 1194 ValLeuAArgIlyThrAlaLeuGlySerValProGlyIlySerIleThrIysGlyIlePro 1213
Db 3631 GTAGTCAGAGGTACCGCA---GGAGCCATACAAAGAAAGATTAATCTGGGGAAATCTCA 3687
Qy 1214 SerThrAArgValProSerAapSerAlaIleThrTyAArgIlySerIleThrHisGlyThr 1233
Db 3688 ACCAGCAAAATTTCAATGAGAGCATTCATCCCTACGGGGGCTGTATCACTCAGGGCACCC 3747

Qy 1234 ProAla-----AapValLeuTyIysGlyThrIleThrAArg 1245
Db 3748 CCGGCTCTTCCCCAGACTGGCATACCAACAGAGCCTTGTGTGAAGGGTCCATTTCAGAG 3807
Qy 1246 IleIleGlyGluAapSerProSerAlaAapAArgGlyAArgGluAapSerLeuProIys 1265
Db 3808 ATGCCCATTTGAAGACGACGAGTCTCT-----GAGAAAGCAGAGAGAGACCTGCATCCAA 3861
Qy 1266 GlyHisValIleTyGlyGlnIlyIysIysGlyHisValLeuSerTyGlyGlnGlyMetSer 1285
Db 3862 GGCCATGTTATTTAAGAAAGCAAAAGTGAACATCTTGTCAATATATATTAAGAAAT 3921
Qy 1286 ValThrGlnCysSerIysGluAapIysSerSerGlyProProIleGluThrAla 1305
Db 3922 GCC-----CGAAGAGGACTAGAGTCCAGAAACAGCTCATGAATCAAGT 3966
Qy 1306 AlaProIysAArgThrTyAapMetMetGlnIlyAArgValGlyAArgAlaIleSer----- 1323
Db 3967 TTA---AAGAGAAGCTATGATCAGTGAAGAAATATTAAGCAAGGATGTCAATGAGG 4023
Qy 1324 -----SerAlaSerIleGlnIlyLeuMetGlyAArgAlaIleProProGluAArg 1339
Db 4024 GAGTCTCTGTATCAGACCGTTAGAGGGCTGATATGCGGACATTAACC-----AGG 4077
Qy 1340 HisSerProHis---HisLeuIysGlnGlnHisIleAArgGlySerIleThrGlnIly 1358
Db 4078 GGGATCTCTCATTTCTGACCTCAAAAGAAAGACTGATTTCTCGGCTCCATATGACGGGG 4137
Qy 1359 IleProAArgSerTyValGluAlaGlnIysAapTyIysAArgAArgGluAlaLysLeuLeu 1378
Db 4138 ACACCAAGAGCAACAACATAAGCTTTGAAGATGGCTT---AAATATCCCAACAAAT 4194
Qy 1379 IysAArgGlnIlyThrProProProProProProSerAArgAapLeuThrGluAlaTyIys 1398
Db 4195 AAAAGGGAAGT-----CCTCCATACGAGCATTTGAAAGTGCAT--- 4236
Qy 1399 ThrGlnAlaLeuGlyProLeuIysLeuIysProAlaHisGlnIlyLeuValAlaThrVal 1418
Db 4237 -----ACCAAGGAACCA---TATGATGGC---ATCACCAACATC 4272
Qy 1419 IysGluAlaGlyAArgSerIleHisGlnIleProAArgGlnGlnLeu----- 1433
Db 4273 AAAGAAATGGGCGTTCCATTCATGATTCAGATTCAGAGCAAGATATTTAATCAGAAAGT 4332
Qy 1434 AArgHisThrProGlnLeuProLeuAlaProAArgProLeuIysGlnIlySerIleThrGln 1453
Db 4333 CGAAAACTCCAGAAAGTGTCCAGAGCACACGCGCATTAATGAGGGTTCATTTCCAG 4392
Qy 1454 GlyThrProLeuIysTyAapThrGlnIlyAaenThrGlySerIlyHisAapVal 1473
Db 4393 GGCACACCAATTAAGTTTACAAAC---TAGGTCAATCTGCCATTCACCAACATATGTC 4449
Qy 1474 AArgSerLeuIleGlySerProGlyAArgThrPheProProValHisProLeuAapValMet 1493
Db 4450 AATCTCTTATACAGGGGCTTGCAGAAATATCCCGGAAGATCCCTCGTGAATATGG 4509
Qy 1494 AlaAap---AlaAArgAlaLeuGlnIlyAArgIlyCysTyGlyIly-----GlnSer 1507
Db 4510 CCGAGAACATTAAGGTGTAGAACGGGGAATAATGAGATGTGAACAGCGGAGACC 4569
Qy 1508 LeuIysSerAArgProGlyThrIleSerSerSerGlyIlySerIleAlaArgIlyAlaPro 1527
Db 4570 GTGCGTTCGGGACACGTCAGTGAAGCTTGGCCCTCGCTTCTTATGATCCACA--- 4626
Qy 1528 ValIleValProGlnLeuGlyIysAProAArgIysSerProLeuThrTyGlnAapHisGly 1547
Db 4627 -----CTGCATGAAGCTCCCAAGACAACTGAGCCCTGGGATTTATATGACACCACT 4680
Qy 1548 Ala-----ProPheAlaGlyHisLeuProAArgGlySerProValThrMet 1562
Db 4681 GCACGAGAGACCCCTGTGATGATTAACAAACACCATGTCCAGAGGCTCACCCATG----- 4734

Db 6655 AGGAGTCACGTC-----TCTTCGAGCCCTACGAGCCCATCTCCCAACCCAG--- 6702
Qy 2231 MetThrGluProGluHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlu 2250
Db 6703 ---GTTCCGGTTGTGCATGAGAAACAGACAGCTTCTGCTCTGTCTCAGAGGGGCCA 6759
Qy 2251 GluThrGluProSerArgMetGlySerIleSerProGluLeuLeuHisSerGlnProProAla 2270
Db 6760 GAGCTCTCAGACGAGGAAATGATGCCGCTCACAGGAGATATAGACTACTTCCCTTCA 6819
Qy 2271 PhePheSerIleValThrGluSerAsnSerAlaMetValIleValSerIleValGlnGluIle 2290
Db 6820 TCTTTCACCAAGCTT---GAAATAACATCACCCATGTTAATCAAGAGACAGAGATT 6876
Qy 2291 AsnIleValSerLeuLeuHisSerArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
Db 6877 TTTTCGTAAGTTGAATCTCTCTGCGAGCGTACCTCTGATATGAGAGCTCTCAGCCAGGA 6936
Qy 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
Db 6937 ACTGAGATCTTAACTTATGCGCAGCAGTTACTACGTCAGGCTCAGTTAGCTCAGAGGCCAT 6996
Qy 2331 AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleValGlyValLeu 2350
Db 6997 TCTTTGCTATCCTGCCAGT---AATCTTGCGCTGGAAGACATTATCAGAGAGCTCTC 7053
Qy 2351 MetGlyValTyrArgProGlnTyrGluGlu-----SerProProLeuSer 2364
Db 7054 ATGGGAAGCTTGTATGCAAAAGTTAGAGATCATGGAATGTGATGCTCCAGCCTATGGA 7113
Qy 2365 AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThr 2384
Db 7114 GTA-----GTGCTGTGACTGCCAACACCTCAGTT-----GTGACC 7149
Qy 2385 AlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly-----Lys 2402
Db 7150 AGTGGTGAACAG 7209
Qy 2403 AlaIleValSerGlyArgProSerSerArgIleValAlaValSerProAlaPro-----Gly 2420
Db 7210 CCAAAGCTGATCAGCAAGTCAAAACAGAGAAATCTAAAGTCTCTTAACTTACCTGGCAAGGC 7269
Qy 2421 LeuAlaSerGlyAspArgProSerValSerSerValHisSerGluGlyAsnProAsn 2440
Db 7270 TACTTAGAAG 7299
Qy 2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
Db 7330 AGGCAACAGCCA-----GGGTGGGCTCTGGGAAGACAGGCCCTCTTCACAGGCTCAACT 7383
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7384 CAGTTTCCTTAAACCTCTGACTATGCGGATG-----CTCAGCACTACTCCACCA 7434
Qy 2481 ProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----His 2496
Db 7435 ACACCGATTGTCATGTGCTCTGCGGTGAACCAAGCACTCTCTCAACAGAGAGAGAG 7494
Qy 2497 AlaTyrAspGluProIleProIleProLeuLeuCysSerGlnTyrGluThrLeuSerAspSer 2516
Db 7495 ATCTGGAGGAG 7554
Qy 2517 Glu 2517
Db 7555 GAT 7557

RESULT 3
US-09-144-085-3/c

; Sequence 3, Application US/09144085

; Patent No. 6280999

; GENERAL INFORMATION:

; APPLICANT: Gustafsson, Claes

; APPLICANT: Betlach, Mary C.

; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Alignment Scores:
Pred. No.: 9,66e-24 Length: 33529
Score: 726.50 Matches: 674
Percent Similarity: 32.36% Conservative: 314
Best Local Similarity: 22.08% Mismatches: 1199
Query Match: 5.50% Indels: 873
Gaps: 137

US-09-522-753-5 (1-2517) x US-09-144-085-3 (1-33529)

Qy 17 ProArgTyrProPro---HisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
Db 22686 CCGAGGCGCTCCACCGACACCATCGCTCCCGGCGCTCGACCCCTGCA-----TCAGCC 22633
Qy 36 AspValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerPro 55
Db 22632 GACCTCGAG-----CGCACTAGTCTGCGCTGCCGCA 22597
Qy 56 GlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGlnProGly 75
Db 22596 GCGTCAGCAACCGCGCAATGGCGAGCGCTCCTATGCAAT-----GCCCAAGCA 22540
Qy 76 AsnGluArgSerGln-----GluLeuHis----- 83
Db 22539 GCCCAGCGGCTCAGCCGCCACGATCCGATGAGAGAGAGAGAGAGAGAGAGAGAG 22480
Qy 84 -----LeuArgPro---GluSerHisSerTyrLeuProGluLeuGly 96
Db 22479 CAGCCGGGCTGCGTACTCCGTCCGCTCCGACAAAGCTCTTCTCGCTGCCCGCGGC 22420
Qy 97 LysSer-----GluMetGluPheIleGluSerLysArgProArgLeuGluLeu 113
Db 22419 CGGGAACACCACTCTCTCAACCCAGGTGAGATAGCGCTCCAGCGCTCCGACACTC 22360
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 22359 GTCGAACGCGCAGCAACAGAGGAGTAACTTGTAAGCTCTTCCCATCCGAGCCG 22300
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 22299 CTGGCTGCCCTG-----CCCTGGAACAGCAC 22273
Qy 143 GlyLeuLeuGluProValSerProProSerProHisThrAspProGluLeuGluLeu 162
Db 22272 CGCAAGCTT---CCGCGCTGCTTCCGCGCGCTCCGCTC-----CACCGCGCATC 22225
Qy 163 ValProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAspArgGlu 182
Db 22224 CGGCGCGCGCGAGAGAGAGCGGAGAGAGCTTCACAGCTCCGCGAGCGCTCCGCAAG 22165
Qy 183 IleThrMetValGlnGlnIleSerLysLeuValLysGlnGlnIleGluGlu 202
Db 22164 CACCGAGCGCGGAGACTCGAAGTGCCTCGGTGCAAGCGCGCGCT-----GCG 22117
Qy 203 GluAlaAlaLysProProGluProGlu-----LysProValSerProProIle 219

[illegible][illegible]

[illegible]

QY	1082	-----HisProLeuProLeuGlyLeuHisSerThrAlaAspProValLeuProArg	1098
Db	19061	GATGGCAATGCGCCCTCGGCTCAAGGCTT-----CGACCAAGCTGCTGCTC	1901
QY	1099	ProProThrIleSerAsnProProProLeuIleSerSerAlaValHisProSerValLeu	1118
Db	19013	CTCTCGGCTGCACCGACCAACACCA-----GGACACCGCGCTCCAGGGGGGTG	1896
QY	1119	GLuArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSer	1138
Db	18965	CCGGCA-----GCTCCACAGCTGAAGACTCACCCTGATCATATGTTGGCCACAG	1891
QY	1139	GLuHisAlaValAlaProValGlyProValThrMetGlyLeu-----	1152
Db	18914	CGTGAACGGCGGTCCGCACTGCTCGGCGCTGGCCGGCCGACATGTAGAGCGCCGATC	1885
QY	1153	---ProLeuProMetAspProLysLysLeuAlaProPheSer	1165
Db	18854	TCGCCACGGCTT-----GCCCGTGAAGTCTGGCACTTGACCGCAAG	1881
QY	1166	-----GlyValLys-----GlnGlnGlnLeuSerPro	1174
Db	18812	CGAGTCTCGGTGTGGCGGCTCGCTCGAGCTTACCCGGACCGGAAGCTCGCTCGCC	1875
QY	1175	ArgGlyGlnAlaGlyProProGlnSerLeuGlyValProThrAlaGlnGlnAlaSerVal	1194
Db	18752	CCGTGCGGTGAGGCGCCACGT-----CCGACACAGCAAAAGCGCACAGCA	1870
QY	1195	LeuArgGlyThrAlaLeuGlySerValProGlnGlySerIleThrLysGlyIleProSer	1214
Db	18707	CTCTGCTCG-----CCCTTGAGCTCTCGAGAAAGCGCGCACCTCG	1866
QY	1215	-----ThrArgValProSerAspSerAlaIleThr---TyrArgLysSerIleThrHis	1231
Db	18665	TGTGGAGCGCGGCTCATAC--GCCGGGTGCACGCCATAGCGCTCGGCTGCTCTTC	1860
QY	1232	GlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg-----Ile-I	1247
Db	18606	TCGGTCCCC-----GGCAGCACACCCGACCGGAATGAGTTGTGCT	1856
QY	1247	egLGLuAspSerProSerArgLeuAspArg--GlyArgGluAspSerLeuProLysG	1266
Db	18564	TGGGCGCAAGCTCCAGAGCCTTGTAAGCGCGGCGGT-----	1852
QY	1266	YHisValIleTyrGlnGlyLysGlyHisValLeuSerTyrGlnGlyMetSerVal	1286
Db	18524	-----AGTGAAGCGCGCGCTCACGACGCGCGT-----	1849
QY	1286	lThrGlnCysSerLysGluAspGlyArgSerSerGlyPro-----ProH	1302
Db	18497	-----CGTAAAGCCGGACAGTTCACCTCTCGCGCGCGCG	1846
QY	1302	sgLThrAlaAlaProLysArgThrTyrAspMetMetGlnGlyArgValGlyAlaAl	1322
Db	18459	ACAGGCAACGTGCACAGCTGTCAGAGCGCACAGGATGCGGGGGCTCTGCTCAACA	1840
QY	1322	eser-----SerAlaSerIleGlnGlyLeuMetCysLysArgAlaIlePr	1336
Db	18399	CTCCCGTCGCTGTGCACCAACGCGCGCTCTCGGGGGCTGTCAAGGCTGCTTACA	1834
QY	1336	oProGlu-----ArgHisSerProHisIleLeuLysGlu-----	1347
Db	18339	GCCCCACCTGGCGCGCGCGCGCGCTCCGAGCATCATCACTGCAGCGCGCGCG	1828
QY	1348	-----GlnHisHisIleArgLys-----IleThrG	1357
Db	18279	CGCCCTCGCGCACGACCAAGCGGTGCGGACGCTGACCTCCAGACGCGCGCTGCCA	1822
QY	1357	ngLylleProArgSerTyrValGlnAlaGlnLysArgTyrLeuArgArgLys-----Al	1375
Db	18219	CCGGCGCGCGCGCGCACGACCAAGCTTCCAGCGTCCCGTCCCGGAGGACCAACCTGGC	1816
QY	1375	AlValLeuLeuLys-----ArgGlnGlyThrProPro-----	1385


```

Db      18159 CAAACACCTCATGATCCGAGAGCCAGAGTGTCTCCCGACGAGCAGCCCGCTGTGAA 18100
Qy      1386 -Pro-ProproposerArgAspLeuThiGluAlaTyrlsThr-----GlnAla 1402
Db      18099 GGTGCCCCCTGGCGGTGGGAGGCTGTGGCGGCGGAGACAGGGGTGCTGCTGCT 18040
Qy      1402 euGlyProLeuLysLeu-Lys-----ProAlaHisGluGlyLeu 1414
Db      18039 TAGAGCCCGGCGAGCTCAGGCTGTAGATATACACCGATACGTTAGTGCC 17980
Qy      1415 ValAlaThrValLysGluAlaGlyArgSerLeHisGluIleProArgLysGluLeu 1434
Db      17979 GCTGAACCCGTAAGTCGGGAGCT-----CCAGAGACAGC---CGC 17941
Qy      1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluLysLeuSerLeHisGly 1454
Db      17940 CCGTGGCCGACAGACCCCGGTCTCAGTCCACT-----CGT 17905
Qy      1455 ThrProLeuLysTyzAspThrGlyAlaSerThrGlySerLysLysHisAspValArg 1474
Db      17904 GCCCGGACGTCGACAGCTGCCAGGCTCTGTACAGCTGGCAGAACCCAC---CTTCGT 17848
Qy      1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db      17847 CCGCGTCAGGCTCCCGACACCA-----CCCGCTGGCGCTCCCGCAGCCGCTGTCA 17794
Qy      1495 AspAlaArgAlaLeuGluAlaGlyAla---CysTyrgLysLeuSerLysSerArgProGly 1513
Db      17793 GCGGACATCCGACAGCCGGGTGGCGCTACCTCCACGAAACACACCGTGCCTCTCCA 17734
Qy      1514 ThrAlaSerSerSerGlySerIleAla-Arg-----GlyAlaProVal11 1529
Db      17733 GAGGCTTCACAGCGCTCGGTCCAGGCGACCGTCCGCGAAGTTGCGGACACAGTACT 17674
Qy      1529 eValProGluLeuGlyLysProArgLys-SerProLeuThrTyrgLysAspHisGly- 1547
Db      17673 CCGCGTCAGCGCTGGCGGCGACACCTCTCTCACTCGTCAAGTAAACGAGAGCT 17614
Qy      1548 -----AlaProPheAlaGlyHis- 1553
Db      17613 GCGTCGCTTCGCGCTGAGCGACAGACAGCTTCGCTCTTACTGGGACAGCGCTTCCA 17554
Qy      1554 -----LeuProArgGlySerProValThiMetArgLysProThrProArgLeu 1570
Db      17553 TGGCGCGGTGCGACGCGTACTGACGTTCACTTCGCGAGAACACACTTTCGCG 17494
Qy      1570 IngLysLeuSerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro 1590
Db      17493 TCAGCTCCACCATCAACCGCTCCACCGCTCCACGTCAACCCGACACACCGCT- 17435
Qy      1590 fGGLuIleAlaLysSer----- 1595
Db      17434 GACGTTTCAACGCGCGCTTCCGACAGCGCTCCCGCAAGCAGACGCTCCGACACACC 17375
Qy      1596 --ProHisSerThrValProGluHis-----ProHisPro1 1608
Db      17374 TCCGACACCGGCGGCGTCAACGACATCATGCCCCATCCCGACGCGTCCGACGCG 17315
Qy      1608 LeSerProTyrgLysLeuLysLeuArgLysVal-----SerGlyValAspLeu 1624
Db      17314 TGGCTTCGAGCGGCGCACTACCCGCGCTCTCCGCAAGCAGACGCTCCGACACACC 17255
Qy      1624 yArgSerHisIleProLeuAlaPheAspPro-----T 1635
Db      17254 GCGCGGACACCTCGCGCTGAGCTGT-GGCCACACACCGCGCAGGCTCCAGCCCGCA 17196
Qy      1635 hSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrgTyLeuProArgHis 1655
Db      17195 CCGCCACCGCGGCGGCGACACCGACGCAACAGGCGGCGTGCACGTCAC 17136
Qy      1655 euAlaProAsnProThrTyrgPro-----HisLeuTyrgProTyrgLeuIleArgGlyT 1673

```

```

Db      17135 CCGCTCCAGCAGCGCTGCTCTCCCGGCATCTCCGCGAGACACCGACAGACCA 17076
Qy      1673 yArgAspThrAlaAlaLeuGluAsnArgLysThrIleLeuAsnAspTyrgIleThrSer 1693
Db      17075 GCCGTCCACGCGCGGAGGCTCATCGACGCGTCA----- 17038
Qy      1693 IngLysMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgLys 1713
Db      17037 -----CCGCTCCGCAACCGCGGCTCTCTCCGACGAGCGCCGCGCC 16995
Qy      1713 euSerProArgLysSerSerLeuAlaLeuAsnTyrgAlaAlaGlyProArgGlyIleAla 1733
Db      16994 CATCCCGGCGCACTGGCTCTCT-----GCCGGGAAACACAAACACACTT 16947
Qy      1733 AspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaT 1753
Db      16946 GCTTCGCGACGCGCGTACCGCGG-----ACACCGCGGATGGCGGACCGTCCGA 16893
Qy      1753 hAlaMetAspArgLeuAlaTyrgLeuProThrAlaProGlnProPheSerSerArgHis 1773
Db      16892 CAGCG-----CCGCAAGCGCTCTCTCCGCTCCGACAGCGCTCGC 16854
Qy      1773 eSerSerProLeu-----SerProGlyLysProThrHisLeuThr 1787
Db      16853 CCGAAGCACCGATGGCGCGTGAAGGCGAAGTGGCTCGGTACAGCGCGCTGTCTTCA 16794
Qy      1787 yArgProThrThrThrSerSerSerGlyArgLysArgAspArgAspArgLysArgAspArg 1807
Db      16793 GTCGACACCATCAGCTCCCGGTCTTCGAGGCACTTCGCGGCGCTCCGCGTCCGCG 16734
Qy      1807 AspArgGluArgLysSerIleLeuThrSerThrThrThrValGlnHisAlaProIleT 1827
Db      16733 CCGCACGAGGCTGTCTCGCGCCCGACAGAACGCGGACATGACAAACCTCTCGTCC 16674
Qy      1827 hArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGlyGly- 1846
Db      16673 C-AAGCGCTCGGCATGCTCTCAACGAGTTCGCGGCGGCTCTCAACGCGCTTCCA 16615
Qy      1847 -----SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro- 1862
Db      16614 GAGTGAACATGCGCGTT-CGTCCGCGTATCCGAAACAGACAGCCCGCGCGCGCGC 16556
Qy      1863 -----IleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHis 1879
Db      16555 CCGCGGTGCGCGCGCACGAGCGCGCTTTCAGAGAACAGCCGCTCCCTCCAC 16496
Qy      1880 ---AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
Db      16495 CCAATGTGCGGAGTGGCTGTCCCGGTGACGAGTCCGCGGACACTTCGTGATC 16436
Qy      1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
Db      16435 GACAGCACATGTAGACACACCGGAGACCGCGCGCA-----GCCCGCGG- 16388
Qy      1919 CysProLeuGlyGlyThrLeuAspGlyValTyrgProThrLeuMetGluProValLeuLeu 1938
Db      16387 TGTCCA---GGTTCGACTTCGACGAGCCAAAGTACAGCGCGCTTCGCGTTCGCGCCAG 16331
Qy      1939 ProLysGluAlaProArgValAlaArgProGluArg----- 1950
Db      16330 GCCCGAACCTTCACAGAGCGTCCGAGCTTCATGCGAGTGGCGCCAGTTCGCGCTCC 16271
Qy      1951 -----ProArgAlaAspThrGly-----HisAlaPheLeu 1960
Db      16270 CATTGGCTTCACCGCGTGTGATGTCTTCGGGAGACGCGGACAGCAGAGAGCGCTGT 16211
Qy      1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db      16210 GAGTACACCGCTGCTGGGAGGAGCGTTTCGCGCGCTCAACCTTGCTGCACGCTCT 16151
Qy      1981 ProArgProLeuVal-----ProProValSerGlyHis----- 1991
Db      16150 GGTTCACCGCTACCCACGATACACGCGACAGACAGCTACCGCTGCGCGCGGCTCCG 16091

```

QY 1992 -----AlaThrIleAlaArgThrPro-----AlaIysAsnLeuAla 2003
 Db 16090 ACAGCGGCTTCAGACACGAGATCCCGACCCCTCGGACGACGATGACCGCCGTCACTGGCG 16031
 QY 2003 ----- 2003
 Db 16030 CAGAGAACTCTTGACACGACCGTCCCGTCAAGGCGCTGAGCGGCTGAATCCACGA 15971
 QY 2004 -----ProHisIleAlaSerProAspProProAlaPro 2014
 Db 15970 ATCCCGGGGTGTGTATCATCATCAACCCACCGGTCAAGCCAGCATGCGACTCGCCCT 15911
 QY 2015 ProAla-----SerAlaSerAspProHisArgIleuArgIleuSerIleu 2029
 Db 15910 GAGCGAGGCGGTGCACGCGACGCGACACACGACGACGACGACGACGACGCGCTGCCA 15851
 QY 2030 ProPheSerIleGlnIleuGlnIleuArgSerIleuIleuArgSerIleu 2049
 Db 15850 CCG-----TGATCGGTGGCCCTGCAGCCCGACGCGACGATACG 15815
 QY 2050 ProGlnIleuAlaIleuProValSerProValSer-----SerProSerIleuThrIleAspIle 2068
 Db 15814 CCAACGCGCTGAGACGACGCTGCGCGCTCCCGGTGCCCGCT----- 15770
 QY 2069 GlyLeuProIleuIleuGlnIleuArgIleuSerIleuGlnIleuArgIleu 2088
 Db 15769 AGCGGTTCACACGCGCGACGCTGCGGTCAAGACGCGGTATCGGAC-----CCA 15716
 QY 2089 LysGlnProGlnProValIleuGlnIleuAlaIleuIleuProIleuArgPro 2108
 Db 15715 TCAGACCCAGGTACACCCCGGTGAGCTCTGCTGACGCGCGACGCGCGACGCGACGCC 15656
 QY 2109 LeuProGlnSerIleuProSerSerSerProIleuGlnIleuArgIleuValIleu 2128
 Db 15655 GCTCA-----GTGCTCCGACGCGCTGCTCA-----CGACGACGCGCTGCTGG 15611
 QY 2129 HisGlnArgValIleuIleuAlaGlnHisIleuSerGlnValIleuThrIleAspIle 2148
 Db 15610 GATCCATCGCGCTGCGCTCTCGGAGCG-----ATATCC 15578
 QY 2149 ArgHisIleuProGlnIleuSerIleuProIleuProIleuArgIleuSerPhePro 2168
 Db 15577 CGAAGAAACCGCATGCA-----ACAGGTGATGTCGCGACGAAATCCACCTCGGACG 15521
 QY 2169 AlaSerCysProValIleuAspIleuArgIleuProProSerIleuIleuProPro 2188
 Db 15520 AGGCTCTGCCA-----CGGCTCCG 15500
 QY 2189 AspHisGlnAlaProIleuArgIleuSerProHisSerGlnIleuIleuValIleu 2205
 Db 15499 GATCGGGGTCTGAATCGAAGGTGCATCCACGCGCGAGGAGGCTCGATGCGCATCTC 15440
 QY 2206 SerProGlnProAsnIleuThrSer----- 2213
 Db 15439 TCCCTCTCTCCAGAGACCGCATGCTTCCGGCGTACGAGCGCCCGCGAGCGCGC 15380
 QY 2214 -----ValIleuGlnIleuGlnIleuArgIleuIleuProVal 2225
 Db 15379 ACGCCATCGATACGATGCGATGCGCGCTGACGAGATCGAGTGTGCTGCGACGCC 15320
 QY 2226 SerProGlnIleuIleu-----ThrGlnProGlnIleuSer----- 2237
 Db 15319 TCTGCAACTCTTCATTCACGCGGTCAACGCTCAAGAACGTAAGAGAGCGCGCTGC 15260
 QY 2238 -----ArgSerAlaValIleuProIleuIleuArg-----Asp 2248
 Db 15259 TATTCATGAGATCTCTGTGATCGAATCATTTGCAAGGCTTCAAGCGCGACGCTGCTG 15200
 QY 2249 GlyGlnIleuThrIleuProSerArgMetGlySer----- 2259
 Db 15199 TCGGCGAGATCTTGAATCGAATGGAACCGAAGCTTCGACGAGATGTTTGGACGCG 15140

QY 2260 ---LysSerProGlnIleuThrSerGlnProProIleuPhePheSerIleuThrGlnSer 2278
 Db 15139 CCGCGACGCTCCAAAGCTTGGGAGAGAGCGCTGAAAGACGACGACGACGACGACGACG 15080
 QY 2279 AsnSerAlaMetValIleuSerIleuValGlnIleuAsnIleuValIleuAsn 2298
 Db 15079 CTGAGCATGAACCGCGCAATTCCTCGTCCGACATGAGTCCGCTGACACGAGATTC 15020
 QY 2299 ArgAsnGlnProGlnIleuIleuSerGlnProGlnIleuIleuPheAsnProAla 2318
 Db 15019 CCGCGCTCACCGTTTCCGAATCTTGACAGAGAAC-----TTCCGAGATGCGGTG 14969
 QY 2319 IleThrGln-----ThrGlnIleuMetThrIleuArgSerGlnAlaVal 2332
 Db 14968 GCAGATGATGATGCTGAAAGAACGACGACGACGACGACGACGACGACGACGACG 14909
 QY 2333 GlnIleuHisIleuSerThrAsnMetGlyIleuGlnAlaIleuIleuArgIleuValIleuMetGly 2352
 Db 14908 CGATTGCGGACTTCGACGCGCATG----- 14885
 QY 2353 LysThrAspGlnIleuGlnIleuSerProProIleuSerIleuAsnAlaPheAsnProIleuAsn 2372
 Db 14884 -----AGCGAATCCATTCCGAGCTCCCTCAAGGCGTGGCGGTAGAAATC 14840
 QY 2373 AlaSerAlaSerLeuProAlaIleuMetProIleuThrAlaIleuAspIleuArgSerAspHis 2392
 Db 14839 GCGTCA-----CCGCGCTGAACCCGACGACGCGCGCGACGCTCGCGCGACGACG 14789
 QY 2393 ThrLeuThrSerProGlnIleuGlnIleuValIleuValIleuSerIleuArgProSerSerArg 2412
 Db 14788 TCGATGACGAGCTC-----AGGCGCTCGCGCTCCGCGACGCGCGAGAGGCGC 14741
 QY 2413 LysAlaIleuSerProAlaProGlnIleuAlaSerGlnAsp-----ArgProProSerVal 2430
 Db 14740 TCGCGAGCGTCAACGCTCTCTTCTCGCGAGAGACGCTTGGCGACGCGCGGCGC 14681
 QY 2431 SerSerValHisSerGlnIleu-----AspCys 2439
 Db 14680 ACCAGCGCGGAAAGACGCGCGCGACGCTGCGCGCTGACGCGCGCGCGCGCGCGCTG 14621
 QY 2440 AsnArgArgThrProIleuThrAsnArgValIleuAspArgProSerSerAlaGlySer 2459
 Db 14620 GCGAGATCCAGGCGACGCTGCGACGAGCTGCTTCAAGAGGTGAGAGCGCGCGCTCAGC 14561
 QY 2460 ThrProPheProIleuAsn-----IleuMet 2469
 Db 14560 AGAGCGAGCGCTCGTGCACCGATATCGGACGACGCGCTGCGCGCTGATGCGGACGAT 14501
 QY 2469 CArgLeuGlnAlaGlnValIleuMetAlaSerProPro-----ProGlnIleuProAla-- 2486
 Db 14500 TCGGCTCGCGCGAGGTGCGCTGCTCATCCGACGCGCTTGGCGCGCGCGCGCGCGCG 14441
 QY 2487 -GlySerGlyProIleuAla 2492
 Db 14440 AGGCTGTCGCGCGGAGCC 14422

RESULT 4
 US-08-372-652-9
 Sequence 9, Application US/08372652
 Patent No. 5932699
 GENERAL INFORMATION:
 APPLICANT: Moore, David
 APPLICANT: Choi, Hwang-sik
 TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESSES:
 ADDRESS: Fish & Richardson P.C.
 STREET: 225 Franklin Street, Suite 3100
 CITY: Boston
 STATE: MA
 COUNTRY: USA

ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1922 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-9

Alignment Scores:
Pred. No.: 1.08e-24 Length: 1922
Score: 718.00 Matches: 212
Percent Similarity: 48.81% Conservative: 76
Best Local Similarity: 35.93% Mismatches: 214
Query Match: 5.43% Indels: 88
Gaps: 25

US-09-522-753-5 (1-2517) x US-08-372-652-9 (1-1922)

QY 1975 ProserlyeGlyserlyrserProgluylValgluProvalserProvalser 1985
DB 209 CCAAGTGGAGGCCCGACCTCATGCTCATGTATTCGTGAGCTGGAAAGATAA 268
QY 1986 -----ProProvalserGlyHisAlaThrIleAlaArgThrProAlaValAsn----- 2001
DB 269 GGGCCCTCTCCAAATTCAGATATGAGGAAGACTAAGACCCGAGGAAAGACTACATT 328
QY 2002 LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
DB 329 ACTGCAGCTAATCTTAGACGTGACCATCACCGCAATTGCTCGGCAAGATGCG 388
QY 2022 HisArgGluylrThrGlnSerlyrProPheSerIleGlnGluylrGluylrGluylr 2041
DB 389 AGGGAACGTGGCTCTCAAGATTCAACTCTCT-----AGTACCTTG 430
QY 2042 GlyTyrHisGlySerSerTyrSerProGluylValGluProvalserProvalser 2061
DB 431 TCTTCTCACAAGATATAACGGCTAGTATGATTCATTAAGTATGATCCCGCAGCTCA 490
QY 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluylrLeuAspLysSerHis 2081
DB 491 CCGCACCAACCCGAGAAAGACAGGCGCTATCAGCCAGACATGGTTAAGCAATCAA 550
QY 2082 LeuGluylrGluylrGluylrProLysGlnProGlyProvalLysLeuylrGluylrAla 2101
DB 551 GGAAGAAATGAGTCCACTCGACAGTATGAAGTCCA----- 586
QY 2102 HisLeuProHisLeuArg-----ProLeuProGluSerGlnProSerSer 2116
DB 587 ---CTGCATCATTTATGCTGCCAGCAGAAATCACATCTCCACAGCAAGCCACACTG 643
QY 2117 SerProLeuLeuGlnThr-----AlaProGlyValLysGlyHisGlnAlaValThr 2134
DB 644 CCCCCCATTTCCACAGTCAAGGGAATGGAGCAAGTCCCGAGCAACCATGACTGATCACA 703
QY 2135 LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisIlePheGlnGln 2154

DB 704 CTTGCTGACCATATGCTCAAAATTATCACACAAGATTGTCTAGAAAT-----CAAGTT 757
QY 2155 LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysPro 2172
DB 756 CCCTGGACGCTTCACTTCACTTCAATTCACAACTTCACCATGCTGTTGTATCCACACT 817
QY 2173 ValLeuAspLeuArgArgProPheSerAspLeuTyrLeuProProProAsp----- 2189
DB 818 GTA-----AGAATAAAACCTCAAGCCGCTACAGCCCAAGATCAACAGTCTCAGACT 868
QY 2190 -----HisGlyAlaProAla---ArgGlySerProHis-----SerGlu 2201
DB 869 GTCTTGATCCAGACACAGCTCTTGAAGTCTTCCAGAAATCTGTGATATAATCCCG 928
QY 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyylrGlyylrGly 2221
DB 929 GGAAGACAGCCCGAAATATCTCCAGAGAGACTCATATC-----CATCAGAGCCC 979
QY 2222 IleGluProvalSerProProGluylrMetThrGluProGlyHisSerArgSerAlaVal 2241
DB 980 TATGAGCCCATCTCCCAACCCCAAGCCCTGCTGTG-----CATGAGAACGAGACAGC 1033
QY 2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
DB 1034 ATGTTGCTCTTGTCAAGAGGAGAGTGAACCTGCTGACAAAGAGTATTCATCTCATCA 1093
QY 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
DB 1094 CCAGAGAGATAGTACGTACTGCTTCATCTTCAACAAGCTT---GAAAGACATACACCC 1150
QY 2282 MetValLysSerLysGlyGlnGluLysLeuLysLysLeuAsnThrHisAsnArgAsnGlu 2301
DB 1151 ATGGTTAAATCAAAACAAACAGAAATTTTTCGTAAATTGAACCTCTGCTGAGAGTGAC 1210
QY 2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
DB 1211 TCTGATATGCGACGCTGCTCAGCAGAAACAGATCTTCAATCTGCTCCAGACTTACACA 1270
QY 2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
DB 1271 TCAGGTGACGTAGAGTCAAGAGCCATTTCTTGTGATCCCGCAGT---AACCTTGCT 1327
QY 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrArgGlnThrGluylr----- 2359
DB 1328 CTAGAAAGCATCATCAAGAAAGCTCTCATGGAAAGTTTGATGATTAAGTCAATCAT 1387
QY 2360 -----SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
DB 1388 GGTGTTGTCATGCTCCATCTCTGTG-----GGCATTTATGCTGTGATGCTCAGACCC 1438
QY 1439 TCAGTG-----GTGACGAGACAGCAGGACCGAGAGATGAAGGAGCA 1483
QY 2396 SerProGlyylrGly-----GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
DB 1484 TCACCTCATGACGAGATATGCAACCAAGCGATCAACAACTCAACACGAGGAAGTCT 1543
QY 2415 LysSerProAlaProGlyLeuAla-----SerGlyAspArgProProSerValSerSer 2432
DB 1544 AATATCTCCATTTCTGGGCAAACTATTAGAACTGAAGGCTTTCTTCTGCTCTCT 1603
QY 2433 ValHisSerGluylrAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
DB 1604 GTGCATTCAGAGAGATTAACACAGCAGACACA-----GATGGCATGGGAAGAT 1657
QY 2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLysMetArgLeuGln 2472
DB 1658 CCGCCCTTTCAACAGGTTTCTCACTCACTTCAACCCCTGTGACCACTAGGAGT--- 1714
QY 2473 AlaGlyValMetAlaSerProProProProGlyLysProAlaGlySerGlyProLeu--- 2491

Db 1715 -----CTCAGACGTACACCACTACAGATCGATCGCCCATCTGCATCACC 1765
Qy 2492 ----AlaGlyProHis-----HisAlaTrpAspGluGluProLysProLeuLeuGly 2507
Db 1766 CAGACACTCCACATCAACAGAACCCGATCTGGAGAGGGAGCGCTGCCCTCTCTCA 1825
Qy 2508 SerGlnTrpGluTrpLeuSerAspSerGlu 2517
Db 1826 GCGCAGTATGAGACATGTCTGATGATGAC 1855

RESULT 5
PCT-US95-16311-9
Sequence 9, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1922 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-9

Alignment Scores:
Pred. No.: 1,08e-24 Length: 1922
Score: 718.00 Matches: 212
Percent Similarity: 48.81% Conservative: 76
Best Local Similarity: 35.93% Mismatches: 214
Query Match: 5.43% Indels: 88
Gaps: 25

US-09-522-753-5 (1-2517) x PCT-US95-16311-9 (1-1922)

Qy 1975 ProSerLysGlySerGluProArgProLeuVal----- 1985
Db 209 CCAAGTGGCAAGGCCCGCCTCATGCTGATGATTCTGAGGCTGGAAAGATAAA 268
Qy 1986 -----ProProAlaSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn----- 2001
Db 269 GGGCTCTCTCAAAATTCAGATATGAGGAAGAGCTAAGAACCCGAGGAAAGACTACCAT 328

Qy 2002 LeuAlaProHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
Db 329 ACTGACACTAATCTTACATAGACGTGACATCACCCGGCAAAATGGCTCGCAACAGATCGC 388
Qy 2022 HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeu 2041
Db 389 AGGAAACGTGCTCTCAAAAGTTCCAGACTCTTCT-----AGTAGCTTG 430
Qy 2042 GLYThrHisGlySerSerLysProGluGluValGluProValSerProValSerSer 2061
Db 431 TCTTCTCAGAGTATGAAGCGCTAGTATGATCCATTAAGTATGATGATCCCGCAGCTCA 490
Qy 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHis 2081
Db 491 CTGACACACCCCGAGAAAGCCACAGCGCTATCAGCCAGACATGCTTAAGCCAAATCAA 550
Qy 2082 LeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGluAlaAla 2101
Db 551 GCAGAAATGATGCCACTCGACAGATGAGAGTCCA----- 586
Qy 2102 HisLeuProHisLeuArg-----ProLeuProGluSerGlnProSerSer 2116
Db 587 ---CTGATCATATATGCTGCTCCAGCAGGAATCACCATCTCCACAGAACCCACCATG 643
Qy 2117 SerProLeuLeuGlnThr-----AlaProGlyValLysGlnAsnArgValAlaThr 2134
Db 644 CCCCACATCTCCAGATCAGAGGGAATGCGACAGTGGCCCGACGCCCATCCATCATGATCA 703
Qy 2135 LeuAlaGlnHisIleSerGluValIleThrGlnAspTrpThrArgHisAspGlnGln 2154
Db 704 CTGCTGACACACATCTGTCAAATTAATACACAAAGATTTCTGAAAT-----CAAGTT 757
Qy 2155 LeuSerAlaProLeuProAlaProLeuLysSerPheProLysAla-----SerCysPro 2172
Db 758 CCTCGAGCGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 817
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuLysLeuProProProAsp----- 2189
Db 818 GTA-----AGAACTAAACCTCAGCGCGTACAGCCGACCAATACAGTCTCAGACT 868
Qy 2190 -----HisGlyAlaProAla---ArgGlySerProHis-----SerGlu 2201
Db 869 GTCTTCATCCAGACAGCAGGCTCTAGAGTCTCCAGAAATCTTGATGATAATCCCGG 928
Qy 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuLysGlyGlyLysArgGly 2221
Db 929 GGAAGCAGCGCTGAAATCTCCAGAGAGGATCATATC-----CATCAGAGCCC 979
Qy 2222 IleGluProValSerProProGluGluMetThrGluProGlyHisSerArgSerAlaVal 2241
Db 980 TATGAGCCCATCTCCACACCCCAAGGCCCTGCTGTG-----CATGAGAAAGCAGACAGC 1033
Qy 2242 TyrProLeuLeuLysArgAspGlyGluGlnThrGluProSerArgMetGlyLysSer 2261
Db 1034 ATGTGCTCTTTCACAGAGGAGGTGACCCCTGTCGACCAAGAGATGATCTTCATCA 1093
Qy 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuLeuThrGlnSerAsnSerAla 2281
Db 1094 CAGAGAGATTAAGTACTTGGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1150
Qy 2282 MetValLysSerLysLysGlnGluIleAsnLysValLeuAsnThrHisAsnArgAsnGlu 2301
Db 1151 ATGGTTAAATCAAAAGAAACAGAAATTTTTCGTAATTAATCTCTGCTGAGAGTGAC 1210
Qy 2302 ProGluLysAsnLysSerGlnProGlyLysGluIlePheAsnMetProAlaIleThrGly 2321
Db 1211 TCTGATATGACAGCTCTCAGCCAGGAAACAGATCTTAATCTGACAGAGTATACACA 1270
Qy 2322 ThrGluLysMetThrLysArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMetGly 2341
Db 1271 TCAAGTGCAGTGCACAAAGAACCACTTTCTTGTGATGCCGCCAGT---AACCTTGG 1327
Qy 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysLysArgProLysGlnGlu----- 2359

[illegible]

Db	14437	CACCCAGCCGCTTCAGCGGGTAAGTCCGGGTCCAGCGGAGCAAGTACAGCGCCCGCCCTTC	14318
Qy	64	ArgArgProSerLeuSerLeuSerGluPheGlnProGlyValSerGlnGluValSerGlnGluLeuHis	83
Db	14377	AGATGCCCGGCGGCTCACACCCAC-----TCCAAAGACGGCTCCGCGCAATGCC	14324
Qy	83	gLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIle	103
Db	14323	ACTCGACCTCCGGGC-----	14308
Qy	103	eGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProle	123
Db	14307	-----	14297
Qy	123	uLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlu	143
Db	14296	CCCTCAGGGTGGCGCAACTGTTGGCTGGCGGATGATCAAGCGCTCTTAATCACTGC	14237
Qy	143	LysLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuVal	163
Db	14236	TGCCTCCGCACTCCATCCATCCAGCGCGCGCGGGGTCTCTCCACTCGCGCTCAACAGC	14177
Qy	163	LProProArgLysSerLysGluLeuLeuIleGlnAspMetAspArgValAspArgGlu-	182
Db	14176	TCCGCAAGCA-----CGCTCCCTGGGATCTCCGCTGC	14141
Qy	183	-----IleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnLeuG	201
Db	14140	CTCCCGCTCCATCTCCACAGAGAGCGCCGCTCTCTCCACGACATCAACGACAGCTCG	14081
Qy	201	LugluGluAlaAlaLysProProGluProGluLysProValSerPro-----	216
Db	14080	CCGAGGC-GGGTCTCGGATTCGACGCGCGGCTCTCCACAGCACCGACATTCGCTCAG	14022
Qy	217	-----ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluVal	234
Db	14021	CAGCGCTCCGCGCTCTC-GCTCGAACAAGCTCGCTGTGTAATG-----	13976
Qy	234	snArgLysValAlaGluAlaAlaHisArgTLeuGluGluLeuGluProGlnValGluL	254
Db	13975	-----ACTGACGGGGGGAGC--TTCGGGAGCTCGACAGCAACATCGACAGCTCGAAC	13922
Qy	254	eupProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisValuAsnIleLysIleAsnG	274
Db	13921	TTCGACGCTGAATCCG-----CGCGTCCATC	13895
Qy	274	ImlaMetArgLysLysLeuIle-----LeuTyrPheLysArgArgAsnHisAlaLysLysG	293
Db	13894	ACCGTCAGGACAAATCCCTCTCCAGCTTCACTTCGGGACAGCGCTGTTCGAATGTC	13835
Qy	293	In-----TrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluL	311
Db	13834	AGGAGAGCTTGAACAACAGCGGCTCGAGCTCAGTCCGGTCCGGCTCGCAATCCTCCAC	13775
Qy	311	ysLysValGluArgIleGlu-----AspAsnProA	321
Db	13774	AGTTCTCGAAGCGAGACTCTGTGTCGAGTACGCCAGGGTCACTCAGCAGCTGC	13715
Qy	321	rguArgArgAlaLysGluSerLysValArgGluTyrTyrGlnLysGlnPheProGluIleA	341
Db	13714	GCCAAACATGTCGCGGAAGAACTCGGACGTAACCGAGCCCTCAGCAGCGGATATGG	13655
Qy	341	rgLysGlnArgGluLeuGlnGlnArgMet-----GlnSerArgValGlyGlnArgG	358
Db	13654	GCGAAAGTTGCCAGTAGAGCCCTCGTCTCGGCTGGGTGAGCAGCGCGATGGGAGGCC	13595
Qy	358	LysSerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspG	378
Db	13594	ACGGTATGTCGTCCTCCCGGAGAGCGCCACAAACACGCTCG-----	13531
Qy	378	LysLeuSerGluGlnGluAsnLeuGluLysGln-----MetArgGlnLeuAlaVal	394
Db	13550	--ATAACCGCGACAGCAACATGAACGGGTGGCCCTTTCGCGCGGGGTGCGCTTC	13499

OY	394	alllePProPMeLeuTyrAspAlaaspGlnGlnArgIleLysPheIleAsnMetAsnG	414
Db	13492	ACTTCGGCCCA-----CAGGCCCTTCGG-----	13470
OY	414	LYLeuMetAlaAspProMeLysValTyrLysAspArgGlnValMetAsnMetTrpSerG	434
Db	13470	-----	13470
OY	434	lunGlnLulYsGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGluLeuI	454
Db	13469	-----CCAGAT-----	13464
OY	454	leAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTrpLysLeuThrL	474
Db	13463	-----CTGC	13460
OY	474	YelYsAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgLysSerG	494
Db	13459	CGCTGGATTCACCTCGAAAGAGTTCGGCCCGGGGAGAGCCTGTCCGAGGAGATCC	13400
OY	494	lunGlnGlnGlnGlnGlnGln-----	502
Db	13399	AGGAATCGCGCGCGCGCGCAGACACTCGCGCAGAGCAACmCGAATTCAAACACC	13340
OY	503	-----GlnGlnG	505
Db	13339	TCGTCCTTCAACCACTCGCGCTCGACAGCGCGTAGTCGCGTAGTCACAGCGGACGTCC	13280
OY	505	lunGlnGlnGlnGlnGlnPProMetProArgSerSerGlnGlnGlnLysAspGluLysGluL	525
Db	13279	GGCAATGTATACGGCAGACCTTTG---CGAAGACGCTCGTAGAGCGCGCCCACTCCGA	13223
OY	525	YsGluLysGluLysGluLysGluLysGluLysProGluValAlaGluAsnAspLysGluAspL	545
Db	13222	AGCAGCACGCTCGCGGACCAACCGTCCAGACAATGTATGCACAACGACGACAGAGAGC	13163
OY	545	euleuLysGluLysThrAspAspThrSerGlyGluAsnAspGluLysGluAlaValAla	565
Db	13162	TGCG-CCTGCTCGACCTTCAACGACAGG-----	13132
OY	565	laSerLysGluArgLysThrAlaAsnSerGlnLysArgArgLysGluArgIleThrArgS	585
Db	13131	-----GGCGACGACGCGCCACGCGGAGCTGAATGCGCGGTCACTTCCGCGCG	13080
OY	585	erMetAlaAsnGluAlaAsnSerGluGlnAlaIleThrProGlnGlnSerAlaGluLeuA	605
Db	13079	CA---CCGCGTGCAGACC-TCGAATCATCTCGGCGCCAGCGTCCACACGAGAGAGT	13024
OY	605	laSerMetGluLeuAsnGlnSerSerArgTrpThrGluGluGluMetGluThrAlaLysL	625
Db	13023	GGACCGGCGAT---GGCGAAGCATCGCTGACAGAGTCTCCGCTTCAATCCGGGAAG-	12968
OY	625	YsGluLeuLeuGluHisGluArgAsnTrpSerAlaIleAlaArgMetValGlySerLysT	645
Db	12967	-----GCCGTACGCAAGACTTCATGACGCGCA	12940
OY	645	hYsAlaSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAsp	663
Db	12939	CCAGCTCCGAG-----AAGCTCGCTCCAAAGCAGAAACAT	12904
OY	664	-----GluIleLeuGlnGln-HisLysLeuLysMetGluLysGluArgAsnAlaArg	680
Db	12903	CCAGGGGGCCCTCCACCACGACGATACGCGTGTATCGGTACCGCCGGCTTCA	12844
OY	681	ArgLysLysLysLysAlaProAlaAlaAlaSerGlnGluAlaAla---PheProProVal	699
Db	12843	GGCGATCCAGGAACCAACATCGCTGCTGGCCCAATACACGCGCACCGGTCCCCCAG	12784
OY	700	ValGluAspGluGluMetGluLysArgLysValSerGlyAsnGluGluMetValGlu	719
Db	12783	GCAATCGACGCAACGCGGACGCGGTCTGACTTC-----	12751

[illegible]

```
Db 11809 TCACAGAGCTCCACGAGGCGG---CCTCTCTCCACACGAGCGCGCTTCGCGCAAC 11753
Qy 1002 -----ProGlnLeuGln-----ProGlnLeu 1009
Db 11752 ACCTGCGCGGCGCGCGCCACTCATCAGCGCGCGGAGTGCGAGGACCAACCTTCANA 11693
Qy 1009 rAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPr 1029
Db 11692 CACTCCACCCCAATGCCCGCTGGAAGTACTCTCCACTCCGCGCGCGCTGCTGCGCCCTCC 11633
Qy 1029 oAlAspLysGlnAlaPheAlaAlaGlnAlaGlnLysLeuProGlyAspProProCysTr 1049
Db 11632 TCGGTGAGCAGCATGCA---GCAGTCCCGCCGTCACAGCGCGCGGTAGACACCGGTGTT 11576
Qy 1049 pHisSerGlyLeuProPheProValProProArgLysValLLeuAlaSerProHisAl 1069
Db 11575 C-----CCAGGTCCCGCGCGCATCGGTGACACAGCGCGCGAAGCTCGTGACCTCC 11528
Qy 1069 aProAspProSerAlaPheSerTyAlaProProGly-----ThrlIeSe 1081
Db 11527 GCCAACCACGCGCGCATTCGCGCGCTTCACGTAGTGGGCAAGCTGCTGCTGACCC 11468
Qy 1081 ----- 1081
Db 11467 GCCAGCGCTTGCGCGCGCGCGCTTCGCGATGTGACAGCATGACCGCATGCTGCTCG 11408
Qy 1082 ---HisProLeuProLeuGlyLLeuHisAspThrAlaArgProValLeuProArg----- 1098
Db 11407 GGTTCGACCTCTACCCGCGGTGCTTCGTGCGCTGCGCTCGACCGCGCGCGGTCTCA 11348
Qy 1099 ---ProPro-----ThrlIeSe 1103
Db 11347 TCCATCTCCACCCGCGCGCGGAGCAGTCCCTGGAAGTGGCGCATGCCGCGCTGCGTC 11288
Qy 1103 rAsnProPro-----ProLeuLLeSerSerAlaLysHisAspSerValLeuGlnArgL 1121
Db 11287 ACCACACCGCGCGCGCGCACTTCTTCACAGCGCGCGCGAGGCGGATTCGTG----- 11236
Qy 1121 nLLeGlyAlaLLeSerGlnGlyMetSerValGlnLeuHisValProTySerGlnHisAl 1141
Db 11235 -----GTTGCGAGGAGTCCAGCGGACGTACCGCGCTCCCGCTTCGAAGATCCCC 11186
Qy 1141 aLysAlaPro-----ValGlyProValThrlMetGlyLeuProLeuProMe 1156
Db 11185 AGCATTCGCCACCAACCCACTCCACGCGCGCTCCAGGACAGAGCGCACTGCACTTCGGG 11126
Qy 1156 cAspProLysLysLeuAlaProPheSerGlyValLysGlnGlu-----GlnLeuSe 1173
Db 11125 CCCACCC-----CCATTCCTCCCACTGGTGGCGCCACTGTTGGCTTGGCGCC 11078
Qy 1173 rProArgGlyGlnAlaGlyProProGlnLeuGlyValProThrAlaGlnGlnAlaSe 1193
Db 11077 TCCAACGTGCGGTACGTCAAGCTCCGCTCTCCGACCGCACCGCGGCGCGCTTGGCGGT 11018
Qy 1193 rValLeuAlaGlyL-----ThrAlaLeuGlySerValProGlyLysSerLLeThLysG 1211
Db 11017 CGCTCGGTGCTGCTCGACACACCGGTGAGACCTCCACCGGAGAGCTGGGCGAG 10958
Qy 1211 LylLePro-----SerThrArgValProSerAspSerAlaLLeThTr 1225
Db 10957 CCGTTCATTCACACGACGCTGATGACGCTCCGATCGGTGACGAG-----CGGT 10907
Qy 1225 yArgGlySerLLeThrHisGlyThrProAlaAspValLeuTyLysGlyThrlLeThra 1245
Db 10906 ATCTGGGAGAGCGCGCATTCGGGCGCGGTGAGC-----GGCTTCACGACCC 10859
Qy 1245 rGllLeLLeGlyGlnAspSerProSerArgLeuAspArgGlyArgGlnAsp----- 1261
Db 10858 TGTCTCAGGTAGAGAACATCCGCTCGC-----CGTGGCGCGCGTGAAGAGTGGTG 10805
Qy 1262 -----SerLeuProLysGlyHisValLLeTyG 1271
```

```
Db 10804 CTGAACCTGAGCAAGCCCTCGAACCCTGCGATCTCCCGAGCTCAAGCAAGCTCG 10745
Qy 1271 LuGlyLysLeuGlyHisValLeuSerTyGlnGlyLysMetSerValThrlGlnCysSerL 1291
Db 10744 AAGCGAGCGCG-----CTCTTGTTCATGAACCCCATC 10712
Qy 1291 yGlnAspGlyL-----ArgSerSerSerGlyProProHis----- 1302
Db 10711 GCGTTCACGCTGACGAGGAGGCGAGCGACCAACGCGCCACCGGCGCGTTCTGCAACAGAG 10652
Qy 1303 -----GlnThrAlaAlaPro-----LysArgThrTyAspM 1313
Db 10651 AGACCTTGAAAGAGCGCGCTGGCGCCAGCTTCCTCCGCTGTCAGATGCTCCACAGAC 10592
Qy 1313 eMetGlnGlyArgValGlyArgAlaLLeSerSerAlaSerLLeGlnGlyLeuMetGlyA 1333
Db 10591 TTCTCGAAGGGAAGTTC-----CTGATG----- 10569
Qy 1333 rGAlaLLeProProGlnLysArgHisSerProHisHisLeuLysGlnGlnHis-----H 1350
Db 10568 --CTGTCAGCGCCCGACGCT--GCGTTCACGCACTGCGCGCAGAGTTCGCGGAAGTG 10514
Qy 1350 LeLLeArg-----GlySerLLeThrlGlnGlyLLePro----- 1360
Db 10513 GCATGAGCTGGAACCTGGCGCGGAGAACACAGAGGTTGACGAAGAACCAAGCCATGAGCC 10454
Qy 1361 --ArgSerTyValGlnAlaGlnGlnAsp-----TyLeuAlaG 1373
Db 10453 TCGCTCTCTGCTGACCGGAGACCGCGGATGGCGAGCCACAGACAGCTCTGCGCC 10394
Qy 1373 rGlnAlaLysLeuLeuLysArgGlnGlyThrlPro-----P 1385
Db 10393 GAGTAGCGGACAGCAGAGAGCTGGAAGCGG--CGAGAGACACATGAAGGCGTGGCTC 10335
Qy 1385 roPro-----ProProProSerArgAspLeuThrGlnAlaTyLysThrlGlnAlaLeuG 1403
Db 10334 CTCCGATGAGGACAGTCCGCTGACCGCGCTGCTCATCGGCTGACACCGCAACCGCAC 10279
Qy 1403 LyProLeuLysLeuLysProAlaHisGlnGlyLeuValAlaThrlValLysGlnAlaGlyA 1423
Db 10278 -----GCATCCATGCAACCGCGGT-----GCGAACAAGTGGCGCGCTGG 10239
Qy 1423 rGSerLLeHisGlnLLeProArgGlnGlnLysLeuArgHisThrlProGlnLeuProLeu--- 1441
Db 10238 CTGTGCGGTGAGTACTCCAGGTGAGAGGCGCGCGACAGCACTGCTGCGGCACCAAGTC 10179
Qy 1442 --AlaProArgProLeuLysGlnGlySer---LLeThrlGlnGlyThrlProLeuLysTyA 1460
Db 10178 GAGCTGCGCTTCAGACAGCGCGCGCTCAGCCATTCGCGCTCCACACCGCGTATGCCG 10119
Qy 1460 apThrGlyAlaSerThrlThrlGlySerLysLysHisAspValArgSerLeuLLeGlySerP 1480
Db 10118 GATCTGACGCGGACAGTCCGCGACCCCGGAGACCGGTCTTGA-----GGAAGGC 10068
Qy 1480 roGlyArgThrlProProValHisProLeuAspValMetAlaAspAlaGlnAlaLeuG 1500
Db 10067 CCGATAGAGCGCGCCACTCCCGGACCA-----GCAAGCCGAGGAGGACGAGCATC 10017
Qy 1500 LuArgAlaCysTyThrGlnGlnLeu-----LysSerArgProGlyThrlAlaSerSers 1518
Db 10016 CGAGACGATGTATGACAGACACAGACACAGGCGGTGACAGTCCGCGAGACCGCGACAG 9957
Qy 1518 er-----GlyGly-----SerL 1522
Db 9956 CCGCAGCGGACACAGGCGCGCGCTCCAGATGAACCGCGCGAGGCTGTTCGATCCG 9897
Qy 1522 LeAlaArgGlyAlaProValLLeValPro-----G 1532
Db 9896 ACGCGCGCGGAATTCGCGCTCCGCTCCACTTCGCGAGCGGTCTCAAGTCTCGAAGGC 9837
Qy 1532 LuLeuGlyLysPro-----ArgGln 1539
Db 9836 GAGCTGAGCTCATAGCGCGATGATGACTGACAGGCTGCGCTCGTGGAGGACAA 9777
```


QY	1539	erProLeuThrTyrgluAmpHisglialaProPhaIaaglyHisleuProArglySerP	1559
		::: ::: ::: :::	
Db	9776	CGTCGTGGCGCA---ACGACTCATGGCGCGGCCAC-AGCGCGAGAAACGCGCTTCACAG	9721
QY	1559	rovalThMeaArglyuProThrProArglyleuGlnGlySerleuSerSerIysa	1579
		::: ::: ::: ::: :::	
Db	9720	CGCTACACTCGAGCGAGCGCCTTGATCCGAG-GACGATGGGAGCATGTACTGGCACTG	9662
QY	1579	laSerGlnaSPArglyleuThrSerThrProArglyIle-----AlaI	1594
		::: ::: ::: ::: :::	
Db	9661	CCCGGATGGAG-----CTGGCGATGAACACAGACCGCTGTGGCGAAGACAGCGGC	9608
QY	1594	ySerProHisSerThrValProGlnHisProHis-----	1606
		::: ::: ::: ::: :::	
Db	9607	AATGGCCCATCGAGGTAAACCGGGTCAACGCTGGCACCTCGCGCCGCTTCCGC	9548
QY	1607	-----ProIleSerProTyrglnHis-----	1613
		::: ::: ::: ::: :::	
Db	9547	GCCGACTCTTGATTCGACGCGCGAGCCCTTGAGCACAGGTACCGAGAGCGTGGC	9488
QY	1614	-----LeuLeuAArglyValSerGlyValaSPLeuTyLr-----ArgSerH	1627
		::: ::: ::: ::: :::	
Db	9487	AGGGCACTCCACGCGCGAAGCGAGGATGTGCGGATGAGCTGCGTGGCCAGAAC	9428
QY	1627	leIleProLeuAlaIleAspProThrSerIleProArglyIleProLeuAraIaIa	1647
		::: ::: ::: ::: :::	
Db	9427	GAGTCCCGGCCCAACTCGAAGAGTCAAGCTTC---CGGCTCACCGCTCATGTC---	9375
QY	1647	laAlaTyTyLleuProArgHisleuAlaProAnProThrTyProHisleuTyLrProp	1667
		::: ::: ::: ::: :::	
Db	9374	-----CAGCACTTACACCA-----	9360
QY	1667	roTyLleuIleArgGlyTyLrProAspThrAlaIaLeuGluAsnArgGlnThrIleIleA	1687
		::: ::: ::: ::: :::	
Db	9359	-----GANGATGCTCAGCTTGCTGCTG	9338
QY	1687	snAPTyIleThSerGlnIleMetHisIleAspThrAlaThaIaMeaIaGlnArgA	1707
		::: ::: ::: ::: :::	
Db	9337	GTGAC-----CGTACGGCGC	9323
QY	1707	IaAPMeIleuAArglyLeuSer---ProArg-----GluSerSerleuAlaIeua	1723
		::: ::: ::: ::: :::	
Db	9322	TCGATGTCTACGCGGCTCGGAAGCCCGCGTCGAGGCCAACATGCTTCGGTGC	9263
QY	1723	snTyAlaIaGlyProArgGlyIleIleAspLeuSerGlnValProHisleuProValI	1743
		::: ::: ::: ::: :::	
Db	9262	ACCTTGGCGGTGGCGTACG-----CGGACAGCGCTCCAGACCA-CCA----	9220
QY	1743	euValProProThr---ProGlyThrProAlaThrAlaMeaAspArgLeuAlaTyLleuP	1762
		::: ::: ::: ::: :::	
Db	9219	-----CAACCGACGACCATGTACTCGGACACTTCTGCTTCAGT-----ACTC	9174
QY	1762	roThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyLp	1782
		::: ::: ::: ::: :::	
Db	9173	CTTCAGGCGCCCTCTCTCTCTCCCGGCGCACACGTAAGCT---CAACCGCTTCTC	9117
QY	1782	roThrHisleuThTyLysProThrThrThrSerSerSerGluArglyuArgAspArgAra	1802
		::: ::: ::: ::: :::	
Db	9116	CTGCCCTCTCCCTCAACACCGCTGCG-----	9088
QY	1802	rgluArgAspArgAspArglyuArglyuSerIleuThSerThrThrThrValG	1822
		::: ::: ::: ::: :::	
Db	9087	-----CTTCTTCACTGGCGGCTGCTG	9066
QY	1822	IuHisAlaProIleTyrArgProGlyThrglnInsSerGlySerSerGlySerSerg	1842
		::: ::: ::: ::: :::	
Db	9065	CTTCAGCGCCTCTTCACCTCCCACTCA-----TCCGTAACCCCTCACTTCAAC	9012
QY	1842	IyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisIleSerP	1862
		::: ::: ::: ::: :::	
Db	9011	CTGCTCTCTCCGACGCCCAAGAACTCCAGAGCCACCTCTCTGCCACCTTCACTTCTC	8952

QY	1862	rolIesEr-----	-----ProAgtThrGlnAspAlaLeuGlnIAspProS	1876
Db	8951	CCCCGTCGGGTACACCCGCTCCCTCCCTCCCAACGGTACGGACGAAACGCTCGCGGT	-----	8892
QY	1876	eVAlleuHnIAsvThrGlyMetLysGlyLeIleThAlaValGluProSerLysProT	1896	
Db	8891	CCACTCCGGCCCTCC-----	-----CCACGTACCCCTCCGACGCTTC	8853
QY	1896	hVAlleuAspSerThrSerThrSerSerPro-----	-----ValArgProA	1910
Db	8852	TCCCTCCACGTACACTCTCCCTCCACCCCAACGGACCGGACCGGACGATTCCTGGTCCAG	-----	8793
QY	1910	lAlAlThr-----	-----PheProProAlaThrHisAcys-----	ProLeuG 1922
Db	8792	CACGTACACCCGCGATTTTCCACCGGCTTCCATGGGCAACGTCAACCCCACTG----	-----	8737
QY	1922	LysGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluA	1942	
Db	8736	-----	-----CTCCTCTGGCAACCGCGTCACTGTCTGAAAGTGTG	8700
QY	1942	lAProArgValAlaArg-ProGluAspProArgAlaAspThrGlyHisAlaPheLeuAla	1962	
Db	8699	CCCTCTGTGGCGCGGTACCGGTTGAGAGAGATGCCCCCACTCTCTCACTG-----	CTCCCT 8643	
QY	1962	LysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluPro	1981	
Db	8642	CACCGCGCCCAACCGCACAGCATGCTCTTC-CCGCGACAGCTTGGCTCACTCACTCAAGC	8584	
QY	1982	ArgProLeuValProProValSerGlyHisAlaThrIle-----	AlaArgThrProAlaLys 2000	
Db	8583	CTTCGGCTGGTACCGCTGCATCTGCTCGAAGAGGCTGCGGTCAAGCACAGCATCTCA	8524	
QY	2001	AsnLeuAlaProHisAlaSerProAspProAlaProAlaProAla-----	SerAlaSer 2019	
Db	8523	CCCTGCTCTTTCAGCTCCGCCCCCAAGCTCTCAACTCCAGGGCGCTTCGGGGTACA	8464	
QY	2020	AspProHisArgGlyLysThrGlnSerLysPheProPheSerIleGlnGluLeuLeuArg	2039	
Db	8463	CCACAGCTTCCCGCGGACAGAGCGGCGCC-----	8411	
QY	2040	SerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProVal	2059	
Db	8430	-----	-----ATACCTCCAAAGCTCACCGGT	8410
QY	2060	SerSerProSerLeu-ThrHisAspLysGlyLeuProLysHisLeuGlnGluLeuAspLys	2079	
Db	8409	CGAAGCCGCTGGCGCCAA-----	TTGCAGCCACACTCTTCGCGCCGGAAGT 8362	
QY	2079	sSerHisLeuGlnGlyLeuLeuArgPro-----	-----LysGlnProGlyProVa 2094	
Db	8361	CCGCGACCGCGCGCCCAACACACAGCGCGGTACCGGACCGTGGGACCGGCAACCCCT	8302	
QY	2094	LysLeuGlnGlyLeuAlaHisLeuProHisLeuArgProLeuProGluSerGlnPro	2114	
Db	8301	TCCGATTTCCCGTCTCCGACCTAAACAATCAAGTACGCGAGGTGGCTCTCTCTAC--	8244	
QY	2114	oSerSerSerProLeuGlnThrAlaProGly-----	-----ValLys 2127	
Db	8243	-TGGGCTCCGGGGTGTGTCTGTGGCTGGCGGAGATCTGCGGTCCCACTCGGTATCCA	8185	
QY	2127	sgLysHisGlnArgValAlaThrLeuAlaGlnHisAlaSerGluValIleThrGlnAspTyr	2147	
Db	8184	CGCACACACACAACTCGGTACCGGACCGCACTTCGCGCAGCTTCTCGCGGACCA	8125	
QY	2147	rThArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePyr	2167	
Db	8124	GCAGCGCGCACCC-CCGCTCCCGCTTCATCCACCCCAAGCGGTCCA-----	GC 8078	
QY	2167	oGlyAlaSerCysArgProAlaLeuAspLysArgArgProSerAspLysTyrLeuProPyr	2187	
Db	8077	GGGTAAGCTCCGGGTCAACGGGACGTACAGCCCCCGCTTTCAGAA-----	8032	
QY	2187	oProAspHisGlyValaProAlaArgGlySerProHisSerGlnGlyGlyLysArgSer--	2206	

```

Db      8031 -----TCCGAGCGGCTCACACCACTCCAGAGCGCTCCGCGACAAAT 7985
Qy      2207 -----ProGluProLeuThyThrSerValLeuGluGlyVal 2218
Db      7984 CCCACTCGCACTCGGGGCCCACTCCCATATCCCTCA-----GctGctGC 7940
Qy      2218 yGluAaPrgIyIleGluProValSerProProGluGlyMetThrGluProGlyVHisSerAr 2238
Db      7939 GCCAAGCTGGTTTG-----CTCGGCGATTCAAGCCTCGTAACCTCAC 7898
Qy      2238 gSerAlaVal---TyrProLeuLeuTyrArgSerGlyGluGlnThrGluProSerArgMe 2257
Db      7897 TCTGCGCCCTCGCCACCTCCACTG-----CCACCGCGGCC 7865
Qy      2257 fGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSer----- 2273
Db      7864 GGGGCTCTCTCCAGCTCGGCTTCAACACAGCTCCGACGAGAC-GCCTCCCTCGGGTACTC 7806
Qy      2274 ---LysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGlnIleAsnLys 2292
Db      7805 CGCTGCGCTCCGCTCCACTCTCCACAGAGAGCGCGCTCTCTCCACGACATCAACA 7746
Qy      2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsn----- 2305
Db      7745 CAGCTCAGACAAAGTCCGGCGCTCCGACACACAGATTGCGCGAGAGCTCCGCGCTCG 7686
Qy      2305 ----- 2305
Db      7685 CTCGACCACTCCGCTCCGCGCGAGTCAATCAAGCCCTCGGCGCTGAGTCAAGCGCGCAG 7626
Qy      2306 ---LleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLe 2324
Db      7625 CTGGCATTTCAGCGCCCGGAGCGGCATCATC-----GTCAACGGAGATG----- 7583
Qy      2324 uMetThrTyrArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGluAl 2344
Db      7582 -----GAGTCACCTCGACGATGTAAGGGTGGCGAACCACCAAGCTCGTCCAGCAT 7533
Qy      2344 alleleArgLysAlaLeuMetGlyLysTyrArgGlnThrGluGluSerProProLeuSe 2364
Db      7532 CGGGGAGAGCCCGCATCCGGGGGAGTATC-----TCGAAGACGACCAAGGCTCTC 7482
Qy      2364 rAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu-ProAlaIleMetProIleT 2384
Db      7481 GAACAACGGG-----GGGCCA-----GGGGGAACCTCGCTCACGGGCGACCTCCA----- 7435
Qy      2384 hAlaAlaAsnArgLysArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaL 2404
Db      7434 -----CCAGGGGCGAGCTCTCATATGCTCGCTCGCTCTCGA-----GCCA 7395
Qy      2404 yValSerGlyArgProSerSerArg-----LysAlaLysSerProAlaProGlyLeuA 2422
Db      7394 GGGTTTCAGCC---CCTTCAGCCAGGGGCCCGACGGGGGCTCGACGGGGCGGCCACTCG 7338
Qy      2422 LAserGlyAspArg-----ProProSerValSerSerValHisSerGluGlyA 2438
Db      7337 CACGGGGATGGCTTGATGAACAAGCCCACTGCTCGATGATCCGGCAAGCTCCGGGGG 7278
Qy      2438 sPcYAsnArgArgThrProLeuThrAsnArgValTrrGluAspArgProSerSerAlaG 2458
Db      7277 CCGCGCGAGACGGTCAACCCGAACAGAGAGT-----CGTCCCTCCGGGTATA 7230
Qy      2458 LysSerProPheProTyrAsnProLeuIleMetArgLeu-----GlnAlaG 2474
Db      7229 GCGCCACAGACAGCGCCCGCGCTTGCAACAGCGTCCCTCCGTACGCCCTGCTG 7170
Qy      2474 LysAlaMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyP 2494
Db      7169 GCGCGAAGAGCGCTCCAGATGCCCGCTCTCTCGGCTCATCGGAGCAGCACTTCACC 7110
Qy      2494 ro-----HisHisAla 2497

```

```

Db      7109 CAGGAGCGTCATACATACGCA 7090
RESULT 7
US-09-522-753-5 (1-2517) x US-09-902-540-9597 (1-14462)
; Sequence 9597 Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9597
; LENGTH: 14462
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9597
Alignment Scores:
Pred. No.: 7,98e-23 Length: 14462
Score: 698.00 Matches: 652
Percent Similarity: 32.50% Conservative: 301
Best Local Similarity: 22.24% Mismatches: 1028
Query Match: 5.28% Indels: 964
DB: Gaps: 143
US-09-522-753-5 (1-2517) x US-09-902-540-9597 (1-14462)
Qy      44 HisLysSerArgAspTyrLysLeuHisLeuSerProGlySerIleIleGlnProGlnArg 63
Db      14437 CACCCAGCGCGCTCCAGCGGGTACTCGGCTCAGCGGACGATCAACGCCGCCCGCTTC 14378
Qy      64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer-GlnGluLeuHi 83
Db      14377 AGAATCCCAAGCGCGCTCACCACCCAC-----TCCAAAGACGCGCTCCGCGCAATGCC 14324
Qy      83 sLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheI 103
Db      14323 ACTCGCAGCTCCGAGG----- 14308
Qy      103 eGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerPro 123
Db      14307 -----CCACTCCCAT 14297
Qy      123 uLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrG 143
Db      14296 CCCCTCAGGTGGTGGCGCCCAAGCTTGCTCGCGCATTCACAGCGCTCGTAACCTCACCTGC 14237
Qy      143 yLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuVa 163
Db      14236 TCCCTCGCCCACTCCACTGCCACCGCGCGGCTCTCTCCACCTGCCCTTCAACAGC 14177
Qy      163 LProPro-ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgL 182
Db      14176 TCCGCGCAGGCA-----CGCTCCCTCGGGATCTCCGCTCCG 14141
Qy      183 ---LleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnIleG 201
Db      14140 CTCGCCCTCCACTCTCCACAGAGAGCGCGCTCTCCACCGACATCAACGACACTCG 14081
Qy      201 LngLgluAlaAlaLysProProGluProGluLysProValSerPro----- 216
Db      14080 CCCAGGC-GCGTCTCGGATTCGCGACCGCGCTTCACAGACCGACATCTGCTCAG 14022
Qy      217 -----ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluA 234
Db      14021 CAGCGCTCGCGCTCTTC-GCCTCGAACAGATCGCTGTGTAATG----- 13976

```

QY 234 vnaRgLybVbAlaGluAlaAlaHisbArg11LeuGluGlyLeuGlyProGluValGlu 254
 Db 13975 -----ACTGACGGGGGGGACAC--TTCCGGACGCTGCACGACGACATGACAGGTTCGAC 13922
 QY 254 euRProLeuTyRbAnGlnProSeRbAspThrArg1nTyRHisGluAen1LeuY1LeaG 274
 Db 13921 TTGACGCTCGAATCCG-----CGCTCCATC 13895
 QY 274 lnaLMeTArgLybVbLeu1Le-----LeuTyRPhelybArgArgbAnHisAlaArgLybG 293
 Db 13894 ACCGTACGGACATCCCTGTGACGCTTCATCTCGGACACAGGCGTTCGTGAAGTTC 13835
 QY 293 ln-----TrpLybGlnLybPheCybGlnArgTyRbArgGlnLeuMetGluAlaLeuGlu 311
 Db 13834 AGGAGACTTGAACAAACAGGGGCTGCCAGTCCGAGTCCGCTCCGCTCGATTCCTCACC 13775
 QY 311 ybLybValGluArg1LeGlu-----AsnAnbProA 321
 Db 13774 AGCTTCTCGAACGGAGCTGCTGCGCGTAGCGCCAGGTCACCTCAGCAGCACTGC 13715
 QY 321 rGaRgRgAlaLybGlySerLybValArgGluTyRtyRgLybGlnPheProGlu1Lea 341
 Db 13714 GCCACATGTCCTCCGAAAGAAAGCTCGGACGTACGGAGCCCTCAGCACAGGATTTG 13655
 QY 341 rGlybGlnArgGluLeuGlnGluArgMet-----GlnSerArgValGlyGlnArg 358
 Db 13654 GCCAAGTGGCCGATGAGCCCTCTGCTCGGCTCGCTGCGACCCGCGAGTGGGAACCC 13595
 QY 358 LySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu1Le1LeaSpG 378
 Db 13594 ACCCTCATGTCGTCTGCGCCCGAGAGCCGACACACACCGCTCG-----13551
 QY 378 LyLeuSerGluGlnGluAenLeuGlnLybGln-----MetArgGlnLeuAlaV 394
 Db 13550 --ATAGCCGCGACGACACATGAACGGCGTGCCTTCTGCGCGCGTGCCTCC 13493
 QY 394 a11LeProPheMetLeuTyRbAspAlaAspGlnGlnArg1LeuPhe1LeaMetbAnG 414
 Db 13492 ACTTGGCGCA-----CAGCCCTTCG 13470
 QY 414 LyLeuMetAlaAspProMetLybValTyRlybAspArgGlnValMetAsnMetTrpSerG 434
 Db 13470 -----13470
 QY 434 lJnGlnGlybGluThrPheArgGlybPheMetGlnHisProLybAsnPheGlyLeu 454
 Db 13469 -----CCACAT-----13464
 QY 454 lea1aSerPheLeuGluArgLybThrValAlaGluCybValLeuTyRtyRLeuThrl 474
 Db 13463 -----CTGC 13460
 QY 474 ybLybAnGluAenTyRlybSerLeuValArgArgSerTyRArgArgArgLybLybSerG 494
 Db 13459 CGCTGATTCACACTCGTGAAGGTCTGGCGCGGAGAGAGCTGTCCTCGGGAGAGTTC 13400
 QY 494 lJnGlnGlnGlnGlnGlnGlnGln-----502
 Db 13399 AGGAATCGGGCGCGCGCCGACACCTCGCGCAGTAGGAACCTGCGATTCCAACACC 13340
 QY 503 -----GlnGlnG 505
 Db 13339 TCGCTTCAACACACTCGCGCTCCAGAGCGCTAGTCCGCTGCTGCACGGGACAGCTCC 13280
 QY 505 lJnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 525
 Db 13279 GGGAAATGCTGACGGGACCTTGG---CGAAGACGCGCTGTAAGCGCGCCGCACTCCGA 13223
 QY 525 ybGlybArgGluAlaGlybGlybGluGlnGlybProGluValGluAenAspLybGluAspL 545
 Db 13222 AGCAGCACGCTCGGAGCAACCGTCCGAGACATGTATGTAACAGCAGCAGCAGGAGCG 13163

QY 545 euLeuTybGlybTyRbAspAspThrSerGlybLeuAsnAspGlybLybGluAlaValA 565
 Db 13162 TGGG--CTGCTCGGACTGCTTCAACAGCAGG-----13132
 QY 565 lAserLybGlybArgLybThrAlaAsnSerGlnGlybArgLybGlybArg1LeThrArgS 585
 Db 13131 -----GGCGACAGCGCGCCGACCGGAGCTCGAATGCGGGTGTGACTTCGCGCG 13080
 QY 585 erMeAlaAnGluAlaAsnSerGluAla1LeThrProGlnGlnSerAlaGluLeuA 605
 Db 13079 CA---CCGGTGGCGACAC--TCGACTCATCTCGGGGCCAGCTCCACACGAGAGCT 13024
 QY 605 lAserMetGluLeuAsnGlySerSerArgTrpThrGluGlnGluMetGluThrAlaLybL 625
 Db 13023 GAGCCGGCAT---GGGAAAGCATGTGCTGACAGAGTGTCCGCTTCATCCGGAAG- 12968
 QY 625 ybGlyLeuLeuGluHisGlybArgbAnTrpSerAla1LeAlaArgMetValGlySerLybT 645
 Db 12967 -----GCCGTACCCAGGACTTCAATGACGCGCA 12940
 QY 645 hrValSerGlnCybLybAsnPheTyRbAnTyRlybLybArgGlnAsnLeuAsp---- 663
 Db 12939 CAGCTCCGAG-----AAGCTCGCTCCACAGCAGAAACAT 12904
 QY 664 -----Glu1LeuGlnGln-HisLybLeuLybMetGluLybGluArgbAnAlaArg 680
 Db 12903 CAGGGGGCCCTCCACCCGACGACATAGGTGCGTGTGATCTGCGTACCGCCGCTTCA 12844
 QY 681 ArgLybLybLybLybAlaProAlaAlaAlaSerGluGluAlaAla---PheProProVal 699
 Db 12843 GCCGATCCAGGAACACAGTGTGCTGCGCGAATGACAGCGGACCCGCTCCCGCAG 12784
 QY 700 ValGluAspGluGluMetGluAlaSerGlybValSerGlybAnGluGluMetValGlu 719
 Db 12783 GCATCGGACGACGCGGACGAGCTGACTC-----12751
 QY 720 GluAlaGluAlaLeuHisAlaSerGlybAnGluVal-----ProArgLybGlu 735
 Db 12750 GATCGCTGCTCCACTGCTGGGCAAGAGCCGAAAGTGGCGCTCGCAAGAGGCGC 12691
 QY 736 CybSerGlyProAlaThrValaAsnAnSerSerAspThrGluSer1LeProSerProHis 755
 Db 12690 GTAGCGGACCTTACGC-----CCAGGACCGCAC 12661
 QY 756 ThrGluAlaAlaLybAspThrGlybAnGlybProLybProProAlaThrLeuGlyAla 775
 Db 12660 GCACACGCG-----CCACCATCTCGTGGCAGTAGAGCAAT 12625
 QY 776 AspGlyPro-----ProGlyProProThrProProArgArgThrSer 790
 Db 12624 GACCACCCGACGCGAAGAGTGTCTCCGCGCCACCGCTCCACCCGCAAG-----12571
 QY 791 ArgAlaPro1LeGlu-----ProThrProAlaSerGlu 801
 Db 12570 CGTGGCCCGACGACGCGCCCAAGACTTTCATGCGGAGAAAGCGGCTCTATAGCGCG 12511
 QY 802 AlaThrGly---AlaProThr-----ProPro-----809
 Db 12510 GAGCCGAGCGACGCGGACAGCTTGGCAGCGGAGCGCGCCAGTCCACCTTGCCAT 12451
 QY 810 -----ProAlaProProSerProSerAlaPro-----818
 Db 12450 TGGGCGTCAAGCGGACAGGGGTTCAGACACACAGCGCCGACACATGTGCCAGAA 12391
 QY 818 -----818
 Db 12390 GCGCTGTGAGAAAGCCCTGAGAGCGCGATGTCCGAGCGCGCTGGCCGTACGA 12331
 QY 819 -----ProProVal-----ValProLyb 824
 Db 12330 CGGCGTACCCACACCGCATTTGTCTCCAGGACCTTCTCGCGGACACACACCGCGT 12271
 QY 825 GluGluLybGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGlu 844

```

Db 12270 CCTGACGTCGGGATGCGCTGTGAGACCGCGCACTTCT--CCGGGCTCCACCGGAAAC 12212
QY      : : : : :
QY 845 ProProAlaAlaGluGluLeuAlaValAspThrGlyValGluGluProValIysSer 864
Db 12211 CCT-----CTCACCTTCACTGGAACTCGTCGCGCGGATGAATCTCC 12170
QY 865 Glu-----CyeThrGluGluAlaGluGluGlyProAlaIleGlyValAspAlaGluAla 882
Db 12169 AGTCCGCAATCCCGCAACACCGCACCTTGTCCCGTGGGTACTCTCTCGCTCCAGCC 12110
QY 883 AlaGluAlaThrAlaGluGlyAlaLeuValAspValGlyValGlySerGlyArg 902
Db 12109 TCCATGCTGACGAGCGGTCGCGACGTACCGCTCGCGCTC-----AACTCCGGGCGC 12059
QY 903 AlaThrThrAlaIysSerSerGlyAlaProGluAspSerAspSerSerAlaThrCysSer 922
Db 12058 CATACGTCACCTCGCTGACCTGCGCCCTCTCCACGAACTCTCCCGCCACGCGCCACCC 11999
QY 923 AlaAspGluValAspGluAlaGluGlyValAspValAspArgLeuLeuSer--ProArgPr 942
Db 11998 -----GGCACAGACGAGGCTCCGGTCCAGACGTAA 11966
QY 942 oSerLeuLeuThrProThrGlyAspProArgAlaAspAlaSerProGluIys-----Pr 960
Db 11965 ACCCGGTGTGCGCATCGCGCGCTCCCA-----GGGCGACCGTCCGCGGTGACTCC 11915
QY 960 oLeuAspLeuValGlnLeuLeuGlnArgAlaAlaAlaIleProProIleGlnVal----- 978
Db 11914 GCTGCACGTG-----GCAGCTTCACCTCGCCCGCGCGACACACCCACCGTTC 11870
QY 979 -----ThrIysValHisGluProProArgG1 987
Db 11869 GTCTCCGTGGGCGCGAGTGTGTGACCTCGACCTCGGCGCGCCAGCGCCGCGACCTCTC 11810
QY 987 uAspAlaAlaProThrIysProAlaProAlaProProIle----- 1001
Db 11809 TCCAGACGCTCCACAGAGGCG--CCTCTCTCCACACACGCGCTTCGTGCGCAAC 11753
QY 1002 -----ProGluAspLeuGln-----ProGluSe 1009
Db 11753 ACCTGCGCGGCGCGCGCACTCATCAGCGCGAGGTGCGAGCGCGCAACCTTCAAA 11693
QY 1009 rAspAlaProGluGlnProGlySerSerProArgGlyIysSerArgSerProAlaProPr 1029
Db 11692 CACTCCACCCCATCGCGCTGGAATCTCTCCACTCCGCGCGCTGCTGCTGCTCTCC 11633
QY 1029 oAlaAspIysGluAlaPheAlaAlaGluAlaGlnIysLeuProGlyAspProProCysTr 1049
Db 11632 TCGTCAGACATGCA--GCAGTCCCCCGTCCACAGCGCGCGTGAACAGCACCGGTTC 11576
QY 1049 pThrSerGlyLeuProPheProValProProArgGluValIleValAlaSerProHisAl 1069
Db 11575 C-----CCAGGTCCCGCGCCACCGCTGCACACGCGCGAAAGCTCGTGACTCC 11528
QY 1069 aProAspProSerAlaPheSerTyAlaProProGly----- 1081
Db 11527 GCCAACCCCAAGCGCGCATTCGCCGCTTCACTAGTATGGGCAAGCTCCGTTGTGACC 11468
QY 1081 ----- 1081
Db 11467 GCCAGCCCTTGGGCGCGCGCTTCCGATGTGAACAGACGTACCGCAAGTTGCTCG 11408
QY 1082 ----HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg----- 1098
Db 11407 GGTTCACCTTACCCGCGGTGCTTCGTGCGCTGCGCTGAGCGCGCGCGGTGTCTCA 11348
QY 1099 -----ProPro-----ThrIleSe 1103
Db 11347 TCCATCTCCACCGCGCGCGCGACGTCTCTCGAAAGTCCCGCACGCGCGCTTCGCTC 11288
QY 1103 rAsnProPro-----ProLeuIleSerSerAlaIysHisProSerValleuGluArgG1 1121

```

```

Db 11287 ACCAACCGGCGCGGCACTTCTTCCACGACGCGCGCGAGCGGATTCG----- 11236
QY 1121 nIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGluHisAl 1141
Db 11235 -----GTTCGAGAGGGTCCAGCGGACGTACGGGCTCTCCCGCTTCGAATTCGCC 11186
QY 1141 aLysAlaPro-----ValGlyProValThrMetGlyLeuProLeuProMe 1156
Db 11185 AGCATCGCACACCACTCCACGCGCCGCTCCAGGACAGCGCACCTGCACTTCGCGG 11126
QY 1156 tAspProIysValAlaProPheSerGlyValIysGlnGlu-----GlnLeuSe 1173
Db 11125 CCCACCC-----CCATTCCCTCAACTGTGTGCCACTGTCGCGCTGCGCGCC 11078
QY 1173 rProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSe 1193
Db 11077 TCCACTGGCGGTAGCTAGCTCCCGCTCTCTGACACCGACGCGGGGCGCTCTGGCGTC 11018
QY 1193 rValLeuArgGly-----ThrAlaLeuGlySerValProGlyGlySerIleThrIysG 1211
Db 11017 CGCTCTGCGGTGCTCGACACACCGGTGACACGCTCCACCGGAGAGCTTGGGCGGAG 10958
QY 1211 yIlePro-----SerThrArgValProSerAspSerAlaIleThrT 1225
Db 10957 CGGTTCATTTCCACGACGACCTGATGAGCTCCGATGCTCAGACG-----CGGT 10907
QY 1225 rYArgGlySerIleThrHisGlyThrProAlaAspValLeuTyIysGlyThrIleThrA 1245
Db 10906 ATCTGGAGAGCGCGGATTCGGCGCGCGGTGTGAC-----GGCTTCAGCACCC 10859
QY 1245 rGlyIleIleGlyIysAspSerProSerArgLeuAspArgGlyValArgGluAsp 1261
Db 10858 TGCCTCAGGTAGAACACCATCCGCTCCG-----CGTGGCGCGCGTGAAGAGTGGTG 10805
QY 1262 -----SerLeuProIysGlyHisValIleTyrg 1271
Db 10804 CTGAACCTGAGACAGCCCTCGAAACCCCGTGAAGTCTCCCGAGCTCAACGAAAGCTG 10745
QY 1271 lGlyIysIleGlyValIleValLeuSerTyrgIysGlyMetSerValThrIlyCysSerL 1291
Db 10744 AAGCGAGCGC-----CTCTGTGTCAATCAACCCCATC 10712
QY 1291 yGluAspGly--ArgSerSerSerGlyProProHis----- 1302
Db 10711 GGCTTCACGCTCAGCGAGGAGGACGACAAACGCGCCACCGGCGCTTTCGAACGAAAG 10652
QY 1303 -----GluThrAlaAlaPro-----LysArgThrTyArgPM 1313
Db 10651 AGACCTGGAAGAGCGCGCTGCGCCACAGTTCCGCTCGGTTGCGATGCTCCAGAGC 10592
QY 1313 eMetGluGlyValGlyValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyA 1333
Db 10591 TTCTCGAAGGAAGTTC-----CTGATG----- 10569
QY 1333 rGAlaIleProProGluArgHisSerProHisIleLeuIysGluGlnHis-----H 1350
Db 10568 --CTGTCAGCCCGCAGCGT--GCGTTCACACCACTGCGCGCAGAGTTCGCGAAAGTG 10514
QY 1350 lAlaIleArg-----GlySerIleThrGlnGlyIlePro----- 1360
Db 10513 GCATAGGCTGGAACCTGGGCGGAGAAACCAAGAGTGTGACAAAGAACCCATGAGCCCC 10454
QY 1361 --ArgSerTyValAlaGluAlaGlnIysAp-----TyLeuArgA 1373
Db 10453 TCCGCTCTCTGCTTACCGGAGACCGCGAGTGGCGAGCCACGACGACGTCCTGCGCCC 10394
QY 1373 rGluAlaIysLeuLeuIysArgGlyGlyThrPro-----P 1385
Db 10393 GAGTAGCGGACAGACGAGGAGCTGGAAGCGG--CCAGACAGACATGAAGGCGGTGCTCC 10335
QY 1385 rProPro-----ProProProSerArgAspLeuThrGluAlaTyIysThrGlnAlaLeuG 1403
Db 10334 CTCCGATGGGCGCAGTGCCTGACGCGCTCATGCGCTCATGCGCTGTACACCGCACCG-- 10279

```

QY 1403 lYpRoLeuLybLeuLybProAlaHISgluGluLeuValAlaThrValIyGluAlaGlyA 1423
 Db 10278 -----GCACCCATGACACCGCGCT-----GGACACAGCTCCCGCGCTCGG 10239
 QY 1423 rGSeRlLeHISgluLLeProArgGluGluLeuArgHISThrProGluLeuProLeu---- 1441
 Db 10238 CTTCGCGGTGGGTAGCTCCAGGTGAGGAGCGCCAGACAGCTGCTGGCCACAGTC 10179
 QY 1442 --AlaProArgProLeuLybGluGlySer---LLeThrGlnGlyThrProLeuLybGlyA 1460
 Db 10178 GACCTGCGCTCTCCAGACGCGCGCTGACCATTCGCGTGCACACCGGTAGTCCGC 10119
 QY 1460 sPThrGlyAlaSerThrThrGlySerLybHISAspValArgSerLeuLISerP 1480
 Db 10118 GTATGACACGGGACGCTCCGGCAACCGGGAGACCGCTTCCTCA-----GGAAGGC 10068
 QY 1480 rGGLyArgThrPheProProValHISProLeuAspValMetAlaAspAlaArgAlaLeuG 1500
 Db 10067 CGCATAGAGCGCGCCCACTCGCGGACCA-----GCACGCCACGACAGCCATC 10017
 QY 1500 LuArgAlaCybTyArgGluGlySerLeu-----LySerArgProGlyThrAlaSerSers 1518
 Db 10016 CGAGACGATGTGATGACAGCACAGACAGCGCGGTGACGTCGCGACAGCGCGACAG 9957
 QY 1518 eR-----GlyGly-----SerI 1522
 Db 9956 CCCACGCGGACACAGGGGCGCGGCTCCCAATGAACGGCCGACGGCTGTTCAGATCCG 9897
 QY 1522 LeAlaArgGlyAlaProValLLeValPro-----G 1532
 Db 9896 ACCGCGCGGGAATTGGCTCGCGCTCCACTTCGCGACGGGTGTCAAGTCTCGAAGGC 9837
 QY 1532 LuLeuGlyLybPro-----ArgGlnS 1539
 Db 9836 GACGTGACGCTCATGCGCCGAGATGATGATGACCTGACAGCGCTGCGCTGTGGGAGGCAAA 9777
 QY 1539 ePProLeuThrTyArgLysAspHISgluValProPheAlaGluLISerProArgLySerp 1559
 Db 9776 CGTCGCGCGCA---ACGACTCATGCGCGCCACC-ACGCGGAGAACCGCGCTGCACAG 9721
 QY 1559 roValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerpLybA 1579
 Db 9720 CGGTACATGAGAGAGCGCTCGATCCGAG-GACGATGGGGAAGTTGTATGCTGGACCTG 9662
 QY 1579 lAserGlnAspArgLybLeuThrSerThrProArgGluLISer-----AlaL 1594
 Db 9661 CCCGATGGAG-----CTGTCGATGAACCAAGCCGCTGTGGCGAAGACAGACCGGCG 9608
 QY 1594 yASerProHISserThrValProGlnHISAspProHIS----- 1606
 Db 9607 AATGCGCCCATGGAGGTACCGGGGTACGCGCTGGACGCTCGGCGCGCGCTTCCCGC 9548
 QY 1607 -----ProLISerProTyArgLISer----- 1613
 Db 9547 GCGGACTCCTCGATTCGACGCGGAGCGCTTCGACGACAGGTACGCGGAGAGCGTCCG 9488
 QY 1614 -----LeuLeuArgGlyValSerGlyValAspLeuTy-----ArgSerH 1627
 Db 9487 AGGGGCAACTCCACGCGGAAAGCGAGCGGATGCGCGCATGAGCTGCGTGGCAGACAC 9428
 QY 1627 lAlLeProLeuAlaPheAspProThrSerLLeProArgGlyLLeProLeuAspAlaAla 1647
 Db 9427 GAGTGCCTCCGCAACTCGAAGATCAGGCTC---CGCGCTCACCGCTCGATGCC--- 9375
 QY 1647 lAlaTyTyTyLeuProArgHISLeuAlaProAspProThrTyProHISLeuTyProP 1667
 Db 9374 -----CAGCACTTCACACCA----- 9360
 QY 1667 roTyTyLeuLISerGlyTyTyProAspThrAlaAlaLeuGluAspArgGlnThrLISer 1687
 Db 9359 -----GATGATGCTCAGCTTCGTCTCG 9338

QY 1687 sNAspTyTLeThrSerGlnGluMetHISAsnThrAlaThrAlaMetAlaGlnArgA 1707
 Db 9337 GTGAC-----CGTACGCGGC 9233
 QY 1707 lAspMetLeuArgGlyLeuSer---ProArg-----GluSerSerLeuAlaLeu 1723
 Db 9322 TCGATGTCTGACGGGGCTCGGAAGCGCCGTCGAGGCGCAACAATGCTTCGCGTCCG 9263
 QY 1723 snTyAlaAlaGlyProArgGlyLISerLeuAspLeuSerGlnValProHISLeuProValL 1743
 Db 9262 ACCCTGCGCTGGGCGCTCAG-----CGCAGCGCGCTTCACACCA---CCA--- 9220
 QY 1743 euValProProThr---ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuP 1762
 Db 9219 -----CCACGAGGAGGACCATGATCTCGCAACTTCGCTCAGT-----ACTC 9174
 QY 1762 roThrAlaProGlnProPheSerSerArgHISerSerSerProLeuSerProGlyLyP 1782
 Db 9173 CTTCAGGCGCGCTCCCTCTCTCCGCGCGCACAGCTAGCGCT---CCACCGCTTCTC 9117
 QY 1782 roThrHISLeuThrLybProThrThrThrSerSerSerGluArgGluArgAspArgAspA 1802
 Db 9116 CTGCGCGCTCTCCCTCACACCGCTCGG----- 9088
 QY 1802 rGGLuArgAspArgAspArgGluArgGlyLybSerLISerLeuThrSerThrThrTyAlG 1822
 Db 9087 -----CTTCCTTCACCTCGCGCTCGT 9066
 QY 1822 LuHISAlaProLISerTyArgProGlyTyThrGlnInsSerGlySerSerp 1842
 Db 9065 CTTCAGGCGCTCTCCACCTCCGCAACTCA-----TCGGTATCCCGCTCACTTCAC 9012
 QY 1842 LyGlyGlyGlyGlySerSerSerArgProLISerHISAsnHISAlaHISGlnHISerp 1862
 Db 9011 CTGCTGTCGCCAGCGCCCAAAATCTCCAGACCCCACTCTCTGCACTCACTCTGTC 8952
 QY 1862 roLISer-----ProArgThrGlnAspAlaLeuGlnGlnArgProS 1876
 Db 8951 CCCCGTCGATACCCGCTCCCGCTCCCAACGGTGAAGGACAGAAAGCTCCGCGT 8892
 QY 1876 eValLeuHISAsnThrGlyMetLybGlyLISerThrAlaValGluProSerLybProT 1896
 Db 8891 CCACTCCGCGCTCC-----CCACGTACCCCTCCGCGCGCTC 8853
 QY 1896 hValLeuArgSerThrSerThrSerSerPro-----ValArgProA 1910
 Db 8852 TCTCCGACGTACACTCTCCCTCAACCCCGACCGGACCGGACGATTCCTTGTCTCAG 8793
 QY 1910 lAlaLThr-----PheProAlaIAsnHISGly-----ProLeuG 1922
 Db 8792 CAGGTACACCGCGGTATTTCCACCGGCTTCGAAAGGACCGTCAACCCCACT--- 8737
 QY 1922 LyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuProLybGluA 1942
 Db 8736 -----CTTCCTTCGAGACGCGGTGACGTCGTGAAAGTCTG 8700
 QY 1942 lAspArgValAlaArg-ProGluArgProArgAlaAspThrLybHISAspLeuAla 1961
 Db 8699 CCCCTGTGCGCGCTGACCCGTTGAGAGATGCCCCCACTCTCACT---CTTCCT 8643
 QY 1962 LyAspProAlaArgSerGlyLeuGluProAlaSerSerProSerLybGlySerGluPro 1981
 Db 8642 CACCCGCGGACGAGACAGACGCTCTC---CGCGGACCTTCGCTCACTCACTCAAGC 8584
 QY 1982 ArgProLeuValProProValSerGlyHISAlaThrLISer---AlaArgThrProAlaLyb 2000
 Db 8583 CTTCGCGGTGTGATGCGCTGCTGCTCGAAGAGGCTGCGTGCACAGAGAGAGCTCA 8524
 QY 2001 AsnLeuAlaProHISAlaSerProAspProProAlaProProAla---SerAlaSer 2019
 Db 8523 CCCCTGCTCTTCAGCGCTCCCGGCGGCGCTTCACCTCCAGCGCGCTTCGCGGTACA 8464
 QY 2020 AspProHISArgGluLybThrGlnInsLybProPheSerLISerGlnGluLeuArg 2039

Db 841474 GTTCGCGAAGGATGAGGCGTGGCGGTGACGCGTTGACGAGATGCGAATGCCCTCTGC 841415
Qy 566 SerIySgIyArgIyThrIaAAsenr-----GInGIyARgY--- 578
Db 841414 TCGCGGTGACGCGAATCTCGAAGATGCGTCAATTGCTTCCAGCGCGGAGGAAAGACGCCCA 841355
Qy 579 -----LySgIyARgIleThrIySgSerMetIaAaenIuaIaAsenr 592
Db 841354 TACCAAGTTCCGACATGTGAAGGGGTGTCAGCCGCGAGCAGTCGATGMAACAGGGGCG 841295
Qy 593 Glu-----GluAlaIleThrProGInGIn----- 600
Db 841294 GACCGTTCCGATTCGAGTTCGCGGTGTCACCGCGTGTGACGCCAGCAGTTCGGCGTGCAG 841235
Qy 601 -----SerIaGluLeuAlaSerMetGluLeuAaenIuSerSerArgTyr 615
Db 841234 AAGGAACATTCGCGCTTTCGCGCTTAATTGGGGGTTACTTCGGAGATTCAAGTTCAGGCC 841175
Qy 616 ThrGluGluGluMetGluThrIaA----- 623
Db 841174 TCGTCGAGCACTTCGCAACGCTCTATTGAGGTTCAGGTGATGCTCCGCGCGAGCG 841115
Qy 624 -----LySlySgIyLeuLeuGluHiSgIyARg-----AsnTyrSerIa 636
Db 841114 GTGCGCGCGTTCGAAACCGGAGCGGTC-----GGCGCGCGCGCGCGACGTCGGCGCG 841061
Qy 637 Ile----- 637
Db 841060 GCCAACTCCGCGCGAAGGCGCTCTCATCATCAGCGCCGACCGATTGACACGCTCGCG 841001
Qy 638 AlaArgMetValGlySerIyThrValSerGInCysIySAsnPhetyrPheAsnTyrIys 657
Db 841000 GCCCGCGGTTTGGCGGACGAGATTTCA----- 840971
Qy 658 LySArgGInLeuAAsenrGluIleLeu-----GInGInHis 669
Db 840970 -----TCGAGATGTGACACCGTGTGTCGATGTCACACGATGACGATCAACACAT 840917
Qy 670 LyLeuLeuMetGluIySgIyArgAAsnIaArgIyLySlySlySlySlySlySlySlySlySly 689
Db 840916 CATTCGCGGTTC-GAGCGCGCGAAGCGATGCTGTGAACGCGCGCGCGTGCAGAGCGCGC 840858
Qy 690 AlaSerGluGluIaAaIaPheProProValIaGluAAsenIuMetGluIaAsenGly 709
Db 840857 CGGTACATCAAGCGCTTACGCCGTTCTACACGAGCAGACCGCGCGCGCGCGCGCGC 840798
Qy 710 ValSerIySAsnGluGluMetValGluGluIaGluIaLeuHiSlyAsenGlyAsn 729
Db 840797 GCCGCTGG----- 840789
Qy 730 GluValProArgIyGluCysSer-----GlyProAlaThrValAAsenSer 745
Db 840788 -----GCCGGTGTACCCCGGATTTCCCGGCGTTACCGCTTTACCGCGCTT 840741
Qy 746 SerAspThrGluSerIleProSerProIleThrGluIaAaIa----- 759
Db 840740 GCGCGCGTTCGCGCGTTCGCGATCAACAGCGCGCTCCGCGACGCGCGCGCTTGGCGCC 840681
Qy 760 -----LySAspThrGlyGlnAAsnGly-----ProLySProProAlaThrLeuGlyAl 775
Db 840680 GATGTGAAAGACACCGGACAAACCGGTGCGCGCGCGCGCGCGCTGCATCTCCGAGACC 840621
Qy 775 AAspGlyPro-----ProProGlyProProThrProPro-----ArgArgThrSerAr 791
Db 840620 GAACGCTCCGCACTTCGGGA 840561
Qy 791 GAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAla 811
Db 840560 GCG 840504
Qy 811 aProProSerProSerIaPro-----ProProValValProLySgIyGlu 827
Db 840503 CCGCGCGCGCGCGCGCTTGGCAGATTGTGAACCCCGCTCCGCGCGCGCGCGCGCGGA 840444

Qy 827 SgIuGluGluThrIaIaAaIaProProValGluGluGlyGluGluGluIySProProAl 847
Db 840443 CGCGAAGACAGAGCTCCG 840408
Qy 847 aAlaGluGluLeuAlaValaAspThrGlyLySlySlySlySlySlySlySlySlySlySly 867
Db 840407 GACCACCGCGCGCGGTACG-GCTGGAACCGCG----- 840375
Qy 867 rGluGluAlaGluGluGluProAlaIySlySlySlySlySlySlySlySlySlySlySly 887
Db 840374 -----GTCG 840322
Qy 887 aGluGlyAlaLeuLySAsnIaGluIySlySlySlySlySlySlySlySlySlySlySly 907
Db 840321 CG 840262
Qy 907 sSerSerGlyAlaProGluAAsenrAspSerSerIaThrCysSerAlaAspGluValAs 927
Db 840261 TCGCGCGCGCGCGAAGCG 840202
Qy 927 pGluAlaGluGlyGlyAspIySAsnArgIyLeuLeuSerProArgProSerIeLeuThrPr 947
Db 840201 TCGTTCG 840142
Qy 947 oThrGlyAspProArgIaAAsnIaAsenrProGInIySProLeuAAsenrIySlySly 967
Db 840141 CCG 840085
Qy 967 sGlnArgAlaAlaAlaIleProProIleGluValThrLySValHiSgIuProProArgI 987
Db 840084 -CCCGGTTC-GCGAACAGCCCGCGTTA-----CCGCGCGCGCGCG 840048
Qy 987 uAspAlaIaProThrIySProAlaProProAlaProProProGluAAsenrIySlySly 1007
Db 840047 GCCATTGCCACGACCG 840008
Qy 1007 oGluSerAspAlaProGInIySProGlySerSerProArgIyLySserArgSerProAl 1027
Db 840007 -----GCCCG 839967
Qy 1027 aProProAlaAspIySgIyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1047
Db 839966 CCG 839907
Qy 1047 oCySTrpThrSerGlyLeuProPheProValProProArgGluValIleLySAsenPr 1067
Db 839906 GAGCAAGCGCGCTCCCG 839847
Qy 1067 oHiSAlaProAspProSerAlaPheSerTyrAlaProProGlyHiAsProLeuProLeuG 1087
Db 839846 GCTGCG 839787
Qy 1087 Y-----LeuHiSAspThrIaArgProValLeuProArgProProThrIleSerAsnPr 1105
Db 839786 CCGCGCGTTCAGCG 839727
Qy 1105 oProProLeuIleSerSerAlaIySlySlySlySlySlySlySlySlySlySlySlySly 1125
Db 839726 GCG 839677
Qy 1125 eSerGIn-----GlyMetSerValGlnLeuHiSlyAlaProTyrSerGInIySAl 1141
Db 839676 TCGCGATCAGCGACCGCGCGGTAGCGTCTGGAAGGGGTCTTCAACATTTATCATCGT 839617
Qy 1141 a----- 1141
Db 839616 CTGTGTGAGAGGTGTGACAGGCGAGAGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCG 839557
Qy 1142 -LySAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLySlySly 1161
Db 839556 ACAGCACCGCGCGGTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839497

QY	1161	LAARProPheSerGIYValLYGInGluGluIleuSerProAArgLYGInAlAGLYProPr	1181
Db	839496	TECCCCCGTGGCCGCGTGGCGGATCAACAGCGGGTCCGCGCGCGGCGCCAC	839437
QY	1181	oGluSerLeuGIYValProThrAla---GInGluAseValIleuArgLYThzAlate	1200
Db	839436	C-----GTCGCCGCGCGCGTGGCTGGCGCGCGTTCGCCCGCTGGCCGT	839389
QY	1200	uGIYSerValProGIYGIYSerIleThrLYs-----GIYIleProSerThzArgVa	1217
Db	839388	TECCGATCAGCCCGGGGGCGCCGCGCGCGCGCGCGCGCGTGTCCGACGCTTTAC	839329
QY	1217	1ProSerApsSerAlaIleThzTYzArgGLYSerIleThzIhISGLYThProAlaSpVa	1237
Db	839328	TAATCCCGCGCGCGCGCGCGAGCCGCGCGAGCGAAGAACATGGCCGCGTCC---	839274
QY	1237	1LeuTYrLYSGIYThzIleThzArgIleIleGIYIuApsSerProSerArgLeuApsAr	1257
Db	839273	-----GCCGGCTCCGCGCGCGCGCGCGCGCGCGCGGAAATCCGCGC	839242
QY	1257	gGIYArgGIuApsSerLeuProLYSGIYhISValIleTYrGIuGLYLYs-----	1273
Db	839241	TECGACTAGCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCATCACAGTGGAGATGCCGCGT	839182
QY	1274	----LYSGIYhISValLeuSerTYrGIuGLYLYMetSerValThzGInCYSerLYSGI	1292
Db	839181	TECCGCGCGCGCGCGCGCG-CCCGCGCAGTGGAGTGGCCGACCGCGCGTCCGCGCTCC	839123
QY	1292	uApsGIYArgSerSerSerGIYProProhISGInThzAlaIleProLYsArgThzTYzAs	1312
Db	839122	CTTCCGCGCGCGCAACAGCGCGCGCGCGCGCGCGCGCGCGCTGCCCGCGTGGAC--	839065
QY	1312	pMetMetGIYArgValGIYArgAlaIleSerSerAlaSerIleGIYLYLeuMetGI	1332
Db	839064	----CGCGGTTCGAGATCCGCGCGAGCGCGCGACCCGCGCGCGCGCAACAGATGCCG	839009
QY	1332	YArgAlaIleProProGIuAArgHISerProhIShISLeuLYSGInGInHIShISLeAr	1352
Db	839008	GGTGTTCGCGCGCGCGCGCGCG-CCCGACCGCGCGCGCGCGCGCGTGGAGTGAAG	838955
QY	1352	gGIYSerIleThzGInGLYIleProApsSerTYzValGIuAlaGInGluApsTYrLeuAr	1372
Db	838954	G-----CCCGCGCGCGCGCGCGCGCGCTGCCGCAACAGCCGCGCGCC	838916
QY	1372	gArgGluAlaLYSLeuLeuLYsArgGLuGLYThzProProProProProSerzArgAs	1392
Db	838915	CGCGCGTTACCGCGCGCGCGCGCGCGCG-GGGGCGCGAACCAGCGCGCGCGCG--	838865
QY	1392	pLeuThzGIuAlaTYrLYsThzGInAlaLeuGLYProLeuLYSLeuLYsProAlaHISGL	1412
Db	838864	-----CCGTTCG-----	838859
QY	1412	uGLYLeuValAlaThzValLYSGInuAlaGLYArgSerIleHISGLIleProAArgGLuGI	1432
Db	838858	-----CCGATCAACACCGCGCGCGCGCGCC	838853
QY	1432	uLeuAArgHISThzProGIuLeuProLeuAlaProAArgProLeuLYSGIYSerIleTh	1452
Db	838852	GTTGGCTCCGCGTCCGCGCGCGCGCGCGCGTGGCGCGCGTTCGCCGATC-----	838790
QY	1452	rGInGLYThzProLeu-----LYSTYzAspTh	1461
Db	838789	-AGCGGCGCGCGCGTGGCGCGCACAGAAATGGCGGTGATCGGGGCGACAGCGCGCAC	838731
QY	1461	rGIYLYAseThzThzGInGLYSer-LYsLYhISAspValArgSerIleGLYSerProAG	1481
Db	838730	GGCGGCGCGCTGGCGCGCGCGCATACGCGCTCCGCGCGCGCTCAGGGGCTGCACGAATGG	838671
QY	1481	LYArgThzPheProProValHISProLeuAspValMetAlaSerAlaArgAlaLeuGLuA	1501
Db	838670	GGCGTAAACGCGCGCG-----CTGTG-----	838648
QY	1501	rGAlaCYeTYrGIuGInuSerLeuLYSAspArgProLY-----ThzAlaSerS	1517

D	838647	-----CGTCCGCGCGCTGGTAGGCTTGCGCTTGGCACCAGCAACAGCGCGC	8386
Q	1517	erSergIyGlySerIleAlaArgIyAlaProValIleValProGluLeuGlyIysProA	1517
D	838601	AATGCGTGTGCACACCTCATCGACACCGCGCGGTGCACAAATCCGTCGTGTGACGGCGC-	8385
Q	1537	rgGlnSerProLeuThrTyTrcIuLamPrlsGlyAlaProPheAlaGlyHisLeuPro----	1555
D	838542	-----CCGCCCATTTAGCCGAGCT	8385
Q	1556	-----ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu	1572
D	838523	AATGCGCAGCCCAATCCCGCCCAATCCGTCGCGCGCCGACCAACAGAGCTCCGCGAAAC	8384
Q	1572	IysSerLeuSerSerSerIyAlaSerGlnAspArgIySLeuThrSer-----	1587
D	838463	CATCACCCATGACATCGCATCTCTCCAA-----CCCTTCACGCGCTGGTCCGACTA	8384
Q	1588	-----ThrProArgGluIleAla-----LysS	1595
D	838409	CCAGCGACACACCCGTAGCGGTAGAAATTAATTCAGAAATCAAGATATCGATCCGCG	8383
Q	1595	erProHisSerThrValProGluHisIAsProHisProIleSerProTyTrcIuHisLeuI	1615
D	838349	CACCGCATGACAGCGCTTGCACACGTTTCGACG-----GCCACAGCTGCTGGAGCGC	8382
Q	1615	euaArgIyValSerGlyValAspLeuTyTrArgSerHisIlePro-----LeuAlaP	1632
D	838288	TCACGGGTGCGCGCGCGCTTGCAGAAACGGCGCGCGGTGCGCGGACAGAGCGCGCGC	8382
Q	1632	heaerProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIaTyTrIleuP	1652
D	838238	CGAAACCCGCA-----CCGCCGGAATAGCGCTGTCACAGACGACCGCGC-----	8381
Q	1652	roArgHisLeuAlaProAsnProThrTyTrProHisLeuTyTr-ProproTyLeuIleArg	1671
D	838194	--ACGGGCTTGTGACAAAGCCCTGAGCGCGCGGAGATCTTCCGCAATACCTGCGCGG	8381
Q	1672	GlyTyTrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyTrIleThr	1691
D	838136	CG-GCCCCCATTTGGGGCGCGC-----TCC	8381
Q	1692	SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnAlaAlaAspMetLeuArg	1711
D	838110	GCAGCGCATCTGCATCACACCGCGGAGTGTGACCCCGGAGCCATCGTGCGCCCGG	8380
Q	1712	GlyLeuSerProArgGlyIleSerSerLeuAlaLeuAsnTyTrAlaIaGly-Pro-----Ar	1729
D	838050	CTACGCGCAACCCGTTGAGCCCGTTTGGCCACAGACGACACCCGCGGCGCGCGGTTC	8379
Q	1729	gGlyIleIleAspLeuSerGlnValProHisLeuProValIleuValProProThrProgl	1749
D	837990	GGGGGTGCGCGCGGTGGCGCGGGCGCGCGCGTTCGCCCGGTG-----CCGCGGGGCGCAT	8379
Q	1749	yThrProAlaThrAlaMetAspArgLeuAlaTyTrLeuProThrAlaProGlnProPheSe	1769
D	837933	CAGCGCGCGCGTGGCG-----CCGACGCGCACCGCGCGCC-----	8379
Q	1769	rSerArgHisSerSerSerProLeuSerProGlyIyProThr-His-----LeuT	1786
D	837899	CCGAAGCCACGACGCGGCGGATTCGCGCGGCGCCCAACGACCCACCGCGCGGTGCCGA	8378
Q	1786	hrlYsProThr-----ThrThrSerSerSerGluArgGluArgAspA	1800
D	837840	CCCCCGCAACCCCGCGCGCGCGCGCGCTGAGACACAGACGCGCGCGGTGCCACGCGGCC	8377
Q	1800	rgAspArgGluArgAsp-----ArgAspArgGluArgGlyIysSerIleLeuThrs	1817
D	837780	GCGGTCGCCACCGAGAAACCAAAATTCGCGCGCGCGCGCGCGCTGCGCAAG	8377
Q	1817	erThrThrThrValGlu-----HisAlaProIleThrPArgProGlyIyThrGluLys-Ser	1834

Db 842237 -----CCGCGGCGCCCGCCATTGCCAGCCGCGCCGACAC 842198
 Qy 997 roAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlyS 1017
 Db 842197 CCGCGCGCGCGCG-----GCCCGCGCGCGCGCCCAAGA 842168
 Qy 1017 eGSePProAArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAla 1037
 Db 842167 GCCGACCG-----GCCGTGGCGCGCGCGCGCCGATGCCACCGACCGCGCGCAC 842117
 Qy 1037 laeLwaGlnGlyLysProGlyLysProProCysTrpThrSerGlyLysProPhePro 1057
 Db 842116 CACCGCTTCCGCGCGCGCGCGCGCGAGCAAGCAAGCGCGCTCCGCGCGCGCCCA 842057
 Qy 1057 aLProProAArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSer 1077
 Db 842056 TCCCGCGAGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 841997
 Qy 1077 yAlaProProGlyLysProLeuProLeuGly-----LeuHisAspThrAlaArgPro 1095
 Db 841996 AGATCCCG 841937
 Qy 1095 aLLeuProAArgProProThrIleSerAsnProProLeuLysSerSerAlaLysHisP 1115
 Db 841936 GCCCGCGGTTCG 841877
 Qy 1115 roSerValLeuGluArgGlnIle----- 1122
 Db 841876 CCGGAGTCCGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 841817
 Qy 1123 -----GlyAlaIleSerGlnIlyM 1129
 Db 841816 CGTTACCATATTGATCATCGTCTCGACAGGCGTGCAGGCGAGGTCTGGCGCGG- 841758
 Qy 1129 eSeSerValGlnLysHisValProTyrSerGlnHisAlaLysAlaProValIleProValI 1149
 Db 841757 --GGCGTAAAGCGGTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841700
 Qy 1149 hTMeGlyLeuProLeuProMetAspProLysLysLysLysAlaProPheSerGlyValLys 1169
 Db 841699 CCGGCGTCCG 841652
 Qy 1169 InGlnGlnLeuSerProAArgGlyGlnAlaGlyProProGluSerLeuGlyValProThz 1189
 Db 841651 TCACAACGCGCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841604
 Qy 1189 la---GlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyLysSer 1208
 Db 841603 CCGGCTCTCG 841544
 Qy 1208 leThriLys-----GlyIleProSerThrArgValProSerAspSerAlaIleThrT 1225
 Db 841543 CCGGCGAG 841484
 Qy 1225 yArGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThz 1245
 Db 841483 CCGCGCGAGCGCAAGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841452
 Qy 1245 rgIleIleGlyLysAspSerProSerArgLeuAspArgGlyArgGlyLysAspSerLeuPro 1265
 Db 841451 -----CGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841397
 Qy 1265 yGgLyHisValIleTyrGlnGlyLysLysLysLysValLeuSerTyrGlnGlyGlyMetS 1285
 Db 841396 CCGGCG 841353
 Qy 1285 eValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGlnThz 1305
 Db 841352 -----GGCGTTG 841325
 Qy 1305 laAlaProLysValGlnThrTyrAspMetMetGlnGlyArgValGlyArgAlaIleSerSer 1325
 Db 841324 CCGGACCG-----GCCGCTG 841310

Qy 1325 laSerIleGlnGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisIle 1345
 Db 841309 CCGCGTTCCTCCCGCGCGCGCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841257
 Qy 1345 euLysGlnGlnHisValIleArgGlySerIleThrGlnGlyIleProAArgSerTyrValG 1365
 Db 841256 CCGCGCGGTGGACCGCG-----CGTTCCGAGATTC 841227
 Qy 1365 lyaLaglnGlnAspTyrLeuArgArgGlnAlaLysLeuLysValArgGlnGlyThrPro 1385
 Db 841226 GCCGACCG 841168
 Qy 1385 ro---ProProProProSerArgAspLeuThrGlnAlaTyrLysThrGlnAlaLeu--- 1402
 Db 841167 CCG 841108
 Qy 1403 -----GlyProLeuLysLeuLysAspProAlaHisGlnGlyLeuValAlaThrV 1418
 Db 841107 ACAGCCCG 841070
 Qy 1418 aLysGlnAlaGlyArgSerIleHisGlnLeuProArgGlnGlnLeuArgHisThrPro 1438
 Db 841069 -----CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841018
 Qy 1438 lueLProLeuAlaProAArgProLeuLysGlnLysSerIleThrGlnGlyThrProLeu- 1457
 Db 841017 GCCCGCGTGGTCCG 840982
 Qy 1457 ----- 1457
 Db 840981 CCGCGTTCGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840922
 Qy 1458 --LysTyrAspThrArgGlyAlaSerThrThrLysSer-LysLysHisAspValArgSerLeu 1476
 Db 840921 GCAGCGCGCACAGCG 840862
 Qy 1477 IleGlySerProGlyLysThrPheProProValHisProLeuAspValMetAlaAspAla 1496
 Db 840861 CCGGACGAACTGGCGGTAAACCGCGCG-----CTGTG----- 840826
 Qy 1497 ArgAlaLeuGlnArgAlaCysTyrGlnGlnSerLeuLysSerArgProGly----- 1513
 Db 840825 -----CGCTCGCGCGCGTGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840793
 Qy 1514 ---ThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGlu 1532
 Db 840792 CGAACACCGCGCGCAATCGCTGTGCAGCACTGCAGCAACCGCGCGCGCGCGCG 840733
 Qy 1533 leuGlyLysProArgGlnSerProLeuThrTyrGlnAspHisGlyAlaProPheAlaGly 1552
 Db 840732 TGTTCAGCGCGCG-----CGCGCG 840715
 Qy 1553 HisLeuPro-----ArgGlySerProValThrMetArgGlnProThrPro 1567
 Db 840714 CATTCAGCGAGTAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840655
 Qy 1568 ArgLeuGlnGlnGlySerLeuSerSerSerValAspGlnAlaArgGlyLysLeuThrSer 1587
 Db 840654 GCTCGCGCAACCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 840601
 Qy 1588 -----ThrProAArgGlnAla----- 1593
 Db 840600 TGGTGGCCCACTACAGCGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840541
 Qy 1594 -----LysSerProHisSerThrValProGlnHisHisProHisProIleSerPro 1610
 Db 840540 TATCGGATCGCGACCGCATGACGCGTTCCACAGTTTCGAG-----GGCGCA 840490
 Qy 1611 TyrGlnHisLeuLeuArgGlyValAspGlyValAspLeuTyrArgSerHisIlePro--- 1629
 Db 840489 GCTGCTGACAGCGGTACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840430

QY 1630 -----LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
DB 840429 ACAGGCGCGCGCGCAACCCGCA-----CCGCCCGAGTAGCGGTCGACAGACGCC 840376
QY 1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProPr 1667
DB 840375 GCG-----ACGGGCGCTGTGACAAAGCGCTGAGCGCCGGGAGATCTTCGCGGAG 840328
QY 1667 cYrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLea 1687
DB 840327 ATACCTGGGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840292
QY 1687 nAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAl 1707
DB 840291 -----TCGGCAGCGCATGTGCATCCAGCGGAGTGGTTAGCGCCCGCAAGCGCAG 840242
QY 1707 aAspMetLeuArgGlyLeuSerProArgGluSerSerIleuAlaLeuAsnTyrAlaAla 1727
DB 840241 CG 840182
QY 1727 Y-Pro-----ArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValP 1745
DB 840181 CG 840125
QY 1745 roProThrProGlyThrProAlaThrAlaMetAlaAspArgLeuAlaTyrLeuProThrAlaP 1765
DB 840124 CG 840086
QY 1765 roGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His 1784
DB 840085 CGGCGCGCG-----CGAAGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840032
QY 1785 -----LeuThrIleuProThr-----ThrHisSerSerSerGlu 1795
DB 840031 CG 839972
QY 1796 ArgGluArgAspArgAspArgGluArgAsp-----ArgAspArgGluArgGluArg 1812
DB 839971 TGGCAGCG 839912
QY 1813 SerIleuThrSerThrThrThrValGlu-----HisAlaProIleThrPargProGly 1830
DB 839911 CGCTGCGCAACAGACCGCTGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839852
QY 1831 ThrGluGln-SerSerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySer 1848
DB 839851 TTTCGCGGTCACCG 839792
QY 1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrG 1868
DB 839791 CG 839732
QY 1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLeuGlyIleLeh 1888
DB 839731 CTCCG 839683
QY 1888 rAlaValGluProSerGlyProThrValLeuArgSerThrSerSer-ProVal 1908
DB 839682 -----CCACCG 839645
QY 1908 rgProAlaAlaThrPheProProAlaThrHisCysArgProLeuGlyGlyThrLeuAspGly 1928
DB 839644 CACCG 839607
QY 1928 aTyrProThrLeuMetGluProValLeuLeuProGlyGluAlaProGlyAlaAlaArgP 1948
DB 839606 -----CCGACCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839561
QY 1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
DB 839560 CGCGCGATCCCG 839507
QY 1968 YLeuGluProAlaSerSerProSerSerGlySerGluProArgProLeuValProPro 1988

DB 839506 CC-----CAACCG 839450
QY 1988 aSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaAla 2008
DB 839449 TACCC-----CACAGCGCG-----GACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839396
QY 2008 erProAspProProAlaProProAlaSerAlaSerAspProHisValGlyValThrGln 2028
DB 839395 CGCGCGACCG 839355
QY 2028 erLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSer 2048
DB 839355 ----- 839355
QY 2048 YrSerProGluGluValGluProValSerProValSerSerProSerLeuThrHisAsp 2068
DB 839354 -----CCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839321
QY 2068 YGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluGlyLeuLeuArg 2087
DB 839320 CGGGGGGCGCG-----TTGCGGATCAGCGCGTCCGCGCGCGCGCGCGCGCGCG 839267
QY 2088 ProLysGlnProGlyProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArg 2107
DB 839266 TGAATCG 839211
QY 2108 ProLeuProGluSerGlnProLeuSerSerSerProLeuLeuGlnThrAlaProGlyVal 2127
DB 839210 GCGTTGAAGCGCTCG 839160
QY 2128 GlnHisGlnArgValValThrLeuAlaGlnHisValSerGluValIleThrGlnAsp 2147
DB 839159 -----GTATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839136
QY 2148 ThrArgHisHisProGlnGlnLeuSerAlaPProLeuProAlaProLeuTyrSerPhe 2167
DB 839135 -----CCGCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839085
QY 2168 GlnAlaSerCysProValLeuAspLeuArgArgProPro----- 2180
DB 839084 GCG 839025
QY 2181 -----SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgLysPro 2198
DB 839024 CGGATCAGCAACCG 838965
QY 2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-----Leu 2215
DB 838964 AACCG 838906
QY 2216 GlnGlyGlyLysLysArgGlyIle--GluProValSerProProGluGlyMetThrGluPro 2234
DB 838905 GCG 838846
QY 2235 GlnHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluPro 2254
DB 838845 CAGCAGACCG 838810
QY 2255 Ser-----ArgMetGlySerLysSerProGlyAsnThrSerGlnPro 2268
DB 838809 ACCGTCACCACTAGCG 838750
QY 2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGln 2288
DB 838749 CCG 838717
QY 2289 GlnLeuAsnLysLeuAsnThrHisAsnArgAsnGluProGlyTyrAsnIleSerGln 2308
DB 838716 ---GTCGTTGAGAAACCAAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838660
QY 2309 ProGly-----ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu--- 2324


```

Db      8836 ||||| : : : ||| ||| |||
      CGCCG-----CAACGACGTGACCGCGCATGCAAGCGCACCGACGACGACGA 8789
Qy      184 ThrMetValGluGlnGlnIleSerIyVbLybLybGlnGlnIleGlu---Glu 202
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      8788 ACA-----CGCCGTCACCGTCAGCGCGCGCGCCCTCCACCCCA 8747
Qy      203 GluAlaAlaLeuPro-ProGlu-----ProGluYProValSer-----ProY 217
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8746 CGCATAGGACCGACCGCGCATGACACTGACGAGGGTCCCGGTAGACACATGCCGCG 8687
Qy      217 oProIleGluSerLybVbIleArgSerLeuValGlnIleIleTyTrpGluAsnArgLyb 237
      | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      8686 GAGAGTGTGCGCGCGCTCCCGACGTGGGTCCG-GTAGTCTGTGAGATGCCCTGA 8628
Qy      237 S-----AlaGluAlaAlaIleArgIleLeuGluGluGluGluProGlnValGlu 254
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      8627 AGACACCGGTCCGGTCCGCGCGACGAGTCCGAAAGATCCGGCCCTCCAGCGCT 8568
Qy      254 uPro-----LeuTyAsnGlnPro-----SerAspThrArgI 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8567 CCCACACGCTCTCCACAGACACCGCTGTCGCGGTCCATGCGCGGTGCGCTCCGCGCGG 8508
Qy      265 nTyTrhIseGluAsnIleLybIle-----AsnGlnAlaMetArgLybLybLeuIle 283
      : ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db      8507 ATATCCCGAAAGACTCCGCGTCAACTCGCGCGGACGACGACGACCGCTTCCCGTA 8448
Qy      283 rPheIyAsArgArgAsnIleAlaArgLybGlnTrpLybGlnLybPheCysGlnArgTyAs 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8447 CATAGCTTTGCCCGGACCCCGG-----CTCCGGGTGTAAAGCCCTCGA 8400
Qy      303 pGlnLeuMetGluAlaLeuGluLybLybValGluArgIleGluAsnAsnProArgArg 323
      ||| : : : : : : : : : : : : : : : : : : : : :
Db      8399 CGTCCACGACGACGCGTCCGCGGACGCGGAGAGCGGTGCGGTACCGCGCAC----- 8345
Qy      323 gAlaLybGluSerLybValArgIleTyTrpGluLybGlnPheProGluIleArgLyb 343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8344 -----CAGGTCCACAGGTCTCTCCGCG 8322
Qy      343 nArgIleGluGlnGluArgMetGlnSerArg----- 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8321 ACGCCAC--ACGCCCGGAGTAACGACGACGACATGCCACGATCGGATCGGTACGCT 8265
Qy      354 -ValGly-----GlnArgGlySerGly-----LeuSerMetSerAlaI 366
      ||||| : : : : : : : : : : : : : : : : : : : : :
Db      8264 CGTCCGCGTACGACCGCGCTGCGCGCGCGCGCTCCGCTCGCTGCTCGGACCGA 8205
Qy      366 aArg-----SerGluIseGluValSerGluIleIleArgGlyLeuSe 380
      : : : : : : : : : : : : : : : : : : : : :
Db      8204 ACAGTTCGTCTGTCAGGTGGAGCGAGCGAGGCGGTCTCGGTGTGTGCA----- 8153
Qy      380 rGluGlnGluAsnLeuGluLybGlnMetArgGlnLeu----- 392
      ||| : : : : : : : : : : : : : : : : : : : : :
Db      8152 -----GAGCAGCGGTCTCGGACAGTTCAGCGCGGTGAGCGCGGTACGCC 8109
Qy      393 -----AlaValIlePro-----PrometLeuTyArgAl 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8108 GGTACGACGATTCGAAACGCTAGAGGAGTCAACCGCGCGACCG-----GA 8061
Qy      402 aaArgGlnArgIleLybPheIleAsnMetAsnGlyLeuMetAlaAspPro-----Me 420
      ||| ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db      8060 AGCGCAGCTCCGGGTGACCGCGCTCGG--CGAGCGGTGCGCAGAGATCCGCGCGCGCT 8004
Qy      420 LybValTyTrpLybAspArgGlnValMetAsnMetTrpSerGluGlnGluLybGlnThr 440
      : : : : : : : : : : : : : : : : : : : : :
Db      8003 CCGAGCGCACAG-----CGTAGCAGAGACGACGCGCT 7968
Qy      440 eArgGluLyb----- 443
      ||| : : : : : : : : : : : : : : : : : : : : :
Db      7967 CGCGCGCGACGCGCGCGACGACGACGCGCGCTCGCGCGCTCTCTCGGTCC 7908
Qy      444 -----PheMetGlnHisProLybAsnPheGlyLeuIleAlaSerPheLeuGlu-- 459
      : : : : : : : : : : : : : : : : : : : : :

```


QY 763 GlyGlnAsnGlyProLys-----ProProAlaThrLeuGlnIleValAlaAspGly 777
|||::|
Db 6868 GGGGAGACA-GCGGCCCCGAGCAAGTACGAGACACCGGGGGGAGACGACTCCGGGAGAGT 6810
QY 778 ProProProGlyProProThrProPro-----ArgArgThrSerArgAlaProIle 794
|||::|
Db 6809 CCGCCCAAGCGCGCCGACGCGCGCGCGAGTTCGAGCGCGGAGACCTGCTCCGATGCCACA 6750
QY 795 GlnProThrProAlaSerGlnAlaThr---GlyAlaProThrProProPro----- 810
|||::|
Db 6749 GCTGAGCGCCCGCAGCAAGAGGGGTCTCGTCCGCGCAGCAGCGCGCGCTCCAGCGGTGG 6690
QY 811 -----AlaProProSerProSer----- 816
|||::|
Db 6689 TGATCCAGACGGGGGCTTGGGGTACCGGCTCGCCGACGCGCTGAGACGAGGCGACGGGCT 6630
QY 817 AlaProProProValValProLysGlnGlnLysGlnGlnIleThrAlaAlaAlaProPro 836
|||::|
Db 6629 CCGCGCGCTCCGGCTCGGCGGGGCGAGCCACAGAACT--CGGGTGTGCGGGCGGACCG 6571
QY 837 ValGlnGlnGlnGlnGlnGlnLysProProAlaAlaGlnGlnIleValAlaAspThrGly 856
|||::|
Db 6570 TCACCGGCGCGCGCGCAACAGGTCCCGGCTGCTCCCG-----CCTGTACCGACCGGC 6517
QY 857 LysAlaGlnGlnLysProValLysSerGlnLysThrGlnGlnIleValGlnGlnIleProAlaLys 876
|||::|
Db 6516 ACCGCG-----ACGGTGTGAGCTCGCGCGCGCGCGCGCGCGAGTTCAGCTGGAGTCCG 6463
QY 877 GlyLysAspAlaGlnAlaAlaGlnAlaThrAlaGlnGlnIleValAlaLysValGlnLys 896
|||::|
Db 6462 GGGGCGAGTCCGGAGTCGGCC---GTGTCTCTCGCGGGGCGAGAGACACATGACCG 6406
QY 897 GlnGly---GlySerGlnArgAlaThrThrAla----- 906
|||::|
Db 6405 GCAGGCGCGGGGTTCGCGGCTGACGACGCGCGCTGCCAGCGAGCGGTAAGCGAGTTC 6346
QY 907 -----LysSerSerGlnAlaProGlnAsp 914
|||::|
Db 6345 CCGGCGCTGCTGACCGCTCCCGCTCAAGTGGCCCGAGCGTCCGCGGAGCGCGGGGCG 6286
QY 915 Ser----- 915
|||::|
Db 6285 TCGAGCGCGGTACGCGCGGTGCGAGCGGGGTCCAGCGAGTACGGGTGCGGTGAAGGGG 6226
QY 916 -----AspSerSerAlaThrCys 921
|||::|
Db 6225 TAGGTGGGAGGTGATGCGCGCGCGGTGCTCCGGTGTGAGTCCGCGGCTCCAGG 6166
QY 922 SerAlaAspGlnValAlaAspGlnAlaGlnGlnGlyAspLysAsnArgLeuLeuSerProArg 941
|||::|
Db 6165 CCGGCGCTCCGGGTGAACACCGTCCGAGCGCGCGCTCCGACACTTCGCGCTCCGACGG 6106
QY 942 ProSerLeuLeu-----ThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
|||::|
Db 6105 CCGCGCGGATGCGCGGACACACATATATCACTCCGCGCGCGAGGACACTACCGCG- 6047
QY 959 LysPro-----LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAla 972
|||::|
Db 6046 CATCCCGGTGAGACCCCATGCGGACCCACCTCCACCAACGTTCCGCCACCCGAGACCG 5987
QY 973 IleProProIleGlnValThrLysValHisGlnProProArgLysAlaAla-----Ala 990
|||::|
Db 5986 CACCAACCCCAACCCCTCCGCAACGACCG--CTTCCGCGGATGACGACCCCAATACC 5928
QY 991 ProThrLysProAlaProProAlaPro-----ProProPro 1002
|||::|
Db 5927 CCGGATCCCAACTCCCGCAACGACCG--CTTCCGCGGATGACGACCCCAATACC 5868
QY 1003 GlnAsnLeuGlnProLys-----SerAlaAlaProGlnGlnProGlnLysSerProArg 1020
|||::|
Db 5867 CACCGGCGCGACCCGAGAACTCCAGCAAGTCCCAACCCCGGAACTCCCA--- 5811

QY 1021 GlyLysSerArgSerProAlaProProAlaAspLysGlnAlaAlaGlnIleGln 1040
|||::|
Db 5810 -----ACACCGGCTCATCAACACCGAATGAACCCATGG-----ACACAT 5769
QY 1041 LysLeuProGlyAspProProCysTrpThrSerGlyLeuProProValProProArg 1060
|||::|
Db 5768 CCAACCGCGCGACTCCACCCCATCCCGCATGACGAGGACGACACCGACTCCA----- 5715
QY 1061 GlnValIleLysValAspProHisAlaPro---AspProSerAlaPheSerTyAlaPro 1079
|||::|
Db 5714 -----GCACACCAATCACTCCGACACACACCGCGGCGCTTACCG 5667
QY 1080 Pro-Gly---HisProLeuProLeuGlnLeu-----HisAspThrAlaAsp 1093
|||::|
Db 5666 CCGCAACCGACACCTCCCGCAACCCCTCAACACCCCGGACACACCGACTCCGACG 5607
QY 1093 GProValLeuProArgProProThrIleSerAspPro----- 1105
|||::|
Db 5606 CCCCACCGACACATCCCGCCACCGCGCAACCCCATCAACCCACCGCGCGCA 5547
QY 1106 -----ProProLeuIleSerSerAlaLysHisProSerValLeuGlnArgGlnIleGln 1123
|||::|
Db 5546 CCAACCAACCGACCGGATCAACGACGACACCCCGCCACATACGACGACG----- 5492
QY 1123 ValAlaIleSerGlnGlnMetSerValGlnLeuHisValProTySerGlnHisAlaLysAl 1143
|||::|
Db 5491 -----CACCTCCCGCACCG--ATGATCCCAACACC 5464
QY 1143 AProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPr 1163
|||::|
Db 5463 ACCGACACCTCCA-----CACCCGAGCTTCCAACCCCGGAGCAACACGCGCAC 5416
QY 1163 oPheSerGlnValLysGlnGlnGlnLeuSerProArgGlnGlnAlaGlyPro----- 1180
|||::|
Db 5415 TCCACGCAAAACGCGGAGCTGAGCAACTCCGTCCGACCCAAAC--AACCCACCCACAGA 5357
QY 1181 -----ProGlnSerLeuGlnValProThrAlaGlnAlaSerValLeuArgGlnIle 1198
|||::|
Db 5356 ACCGAAACCAAGAACGACCCCAACCGGACCGGACCGGATTCAGCCCGGACACCGG 5297
QY 1198 rAlaLeuGlnSerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
|||::|
Db 5296 GTCCAGTGCAGCTCCGCGAATCATCCGCGCAACGACCGACCTCCACCTCCCGCAC 5237
QY 1218 oSerAspSer-----AlaIleThrTyArg 1226
|||::|
Db 5236 CATGCAACAACCTCATTCAGACACTCCGCAACACCCCGCGCAACACCCAGC 5177
QY 1226 gGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIle 1246
|||::|
Db 5176 CCCCATCAACACCTGCGGACCTGACCCGTG-----MACAACACCCCAAC- 5128
QY 1246 eIleGlyLysAspSerProSerArgLeuAspArgGlyArgGlyAspSerLeuProLysGln 1266
|||::|
Db 5127 -----CACCGCGGACGCGCGCGGACCGGACGACACCCCGAGACACCTCCCGGACG 5075
QY 1266 yHisValIleTyGlnGlnGlnLysValHisValLeuSerTyGlnGlnGlnMetSerVal 1286
|||::|
Db 5074 AAC----- 5071
QY 1286 ThrGlnCysSerLysGlnAspGlyArgSerSerSerGlyProProHisGlnThrAlaAl 1306
|||::|
Db 5070 -----CCACCAAGACGACGAGCTCCCAAGCCCAACGACGACGACGACGACG 5021
QY 1306 aProLysArgThrTyAspMetMetGlnGlyArgValGlyArgAlaIleSerSerAlaSe 1326
|||::|
Db 5020 A-----CGGTCTCGAAAGAGTCCGCGCTCGAAGC 4991
QY 1326 rIleGlnGlnLysMetGlyArgAlaIleProPro----- 1337
|||::|
Db 4990 CAGCGAATAGCAACATCCCGGAGATCCGACCCGCGCTCACTATCAACCAACG 4931
QY 1338 -----GlnArgHisSerProHisIleLeuLysGlnGlnIleHisIleArgGlySe 1354

Db 4930 CACCGCTGCGCGCCCAAGCCCTCCGACCGCGCGCAACAAACAGCGCGAGCGAGC 4871
 Qy 1354 rtlethrnglnlyleprothrserlytrvalgluablnghuaprrtyrleuabrgl 1374
 Db 4870 AGCCGCGTCCGCGGTACCGGGTCAAGAGCAGCGCCGTC-GATGCTCCAGCCACT 4812
 Qy 1374 ualalyseuleu 1382
 Db 4811 CCGGCGCTCTCTGATGATCAGATGCGGTGCTGCCACTGATCCGGAACAGACAGCAG 4752
 Qy 1382 ythrpro-----ProProProProProSerArg-----As 1392
 Db 4751 CCGCGCGCGCGCGCGGTCCGACCGCGCGCGCACCGCGCGCTCTGCGACAGGCGCA 4692
 Qy 1392 pleuThrGluAlaTyrylyseThrGlnAlaLeuGlyProleuylseuylseuProAlaHlsGl 1412
 Db 4691 CCGTCCCGCAGCAGCAGTCGACGTGCGTCTGGCGCGGTCCGCGTCAAGTTCGCGGCA 4632
 Qy 1412 uGlyLeuValAlaThrVallyseGluAlaGlyAspSerIleHlsGluIleProArgGluGl 1432
 Db 4631 GCGTCCGCGTCCGACGGGCAAG-----CAGCA 4605
 Qy 1432 uLeuArgHlsThrProGluLeuProleuAlaProArgProleuylseuGlySerIleTh 1452
 Db 4604 TCTTGACCAACAC-----CGCGCGCGCGCG-----GGCGCGCTG-- 4571
 Qy 1452 rGlnGlyThrProleuylseTyAspThrGlyAlaSerThrThrglySerIlelyHlsAs 1472
 Db 4570 -----GCTGACCGACGCTTGAGCTTGACGTTGACGTTGACGAGCCCA 4539
 Qy 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHlsProleuAla 1492
 Db 4538 GCCACAGGGGTC-----GCCACCGCGTC----- 4514
 Qy 1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCyStryGluGlySerIlelySerArgPr 1512
 Db 4513 ---ACGCCGTACGTGGCGCAGCAGCGCGCTCGCTCGATCGGATGCCAGCGCGCTGC 4458
 Qy 1512 oGlyThrAla----- 1516
 Db 4457 CCGTACCGTGGCGCTCCACCGCGCTCCAGTCACCGCGCGAGCGCGCGTGGCGAGG 4398
 Qy 1516 rSerSerGlyGlySerIleAlaArgGlyAla-----Pr 1527
 Db 4397 CAGCTCGATGACGCGCTCTCTGCGCGAGCGGTGGAGACGGTCAGCGCGTGTGCGAC 4338
 Qy 1527 oValIleVal--ProGluLeuGlylyseProArgGlnSerProleuThrTyThrAlaHls 1546
 Db 4337 CGTCTGCTTGAACGCACTGCCCGCGCACCGCCAGCACCC----- 4296
 Qy 1547 GluAlaProPheAlaGlyHlsleuProArgGlySerProValThMetArg----- 1563
 Db 4295 GGTGTCCCGCAGCGCGCGCT--CGAAGACCGCTCAAGTCACTACGACGACACCCCT 4239
 Qy 1564 -----GluProThrProArgLeuGlnGlySerLeuSerSerSerIleAlaSerGln 1581
 Db 4238 CGGACCGAGCCGTACCGTGGCATCGCGGAGAGACGACATCGAGCGCGTGGGGGCA 4179
 Qy 1582 AspArgGlyLeuThrSerThrProArg-----GluIleAlaLyseProHlsSerThr 1599
 Db 4178 GCCCGCGCTGCCCGAGAACTCCACGAAATGCGCGGCGCCGACATCCGTTGGCGCGC 4119
 Qy 1600 ValProGluHlsHlsProHlsProIleSerPro----- 1610
 Db 4118 CCGCCAGAGCC-AGCCCGAC-----TCGCCCCGCGCAGCGCGCTGCAAGCGCAGGTGC 4066
 Qy 1611 -----TyrgluHlsleuLeuArgGlyValSerGlyValAspLeuTyArg 1625
 Db 4065 AGCGCCACAGGAGCAGAGGACGCGCGGTGTGACGCTGAC----- 4024
 Qy 1626 SerHlsIleProleuAlaPheAspProThrSerIleProArg-----GlyIleProleu 1643

Db 4023 -----GCCGTCCGCG-----GTCCGAGGCGTTCAGCATCCGCGCG 3985
 Qy 1644 AspAlaAlaAlaTyrylyseThrProArgHlsleuAlaProAspProThrTyrrProHls 1663
 Db 3984 GAGGCGAGCTCGCG-----GTGGTCCCGGTAGACGAGTAACTCCCTGC 3943
 Qy 1664 LeuTyPro-----ProTyLeuIleArgGlyTyPro-- 1674
 Db 3942 ---TATCCGTCGCTCCCTCGTGAAGCTCGGCGCTGATCTCCGTCGCGTGGCGCGAGC 3886
 Qy 1675 AspThrAlaAlaLeuGluAspArgIleHlsIleAlaAspTyryleThrSerGln 1694
 Db 3885 AACACTCCGCGGTCCCGCGCGCGAGTCACTGGGAGACATCCGCGCTCTGAGAGGCC 3826
 Qy 1695 MetHlsAspThrAlaThr-----AlaMetAlaGlnArgAla 1707
 Db 3825 TCCACAGACGCTTCAGACAGCAAGCGCTGCGGCTCATGCGCTGCGCTCGCGCGG 3766
 Qy 1708 AspMetLeuArgGlyLeuSerProArgGlySerSerLeuAlaLeuAspTyryAlaAlaGly 1727
 Db 3765 GATATC-----CCGAAAGAACTCC-----GCGTCGAATCCGCGCGCC-- 3730
 Qy 1728 ProArgGlyIleIleAspLeuSerGlnValProHlsleuProValLeuValProProThr 1747
 Db 3729 -----GAGTCAAGAAACCGCGCTCCCGACATAGGTCTTGCGCGGCGTC 3685
 Qy 1748 ProGly-----ThrProAlaThrAlaMetAspArgLeuAla-----Tyr 1760
 Db 3684 CCGGCGTGGGGTTCGTACAGCCCGCTCGAGGTCCAGCGCGGTGCGTGGGAAAGACAG 3625
 Qy 1761 LeuProThrAlaProGlnProPheSerSerArgHlsSerSerSerProleuSerProGly 1780
 Db 3624 ATGGCGTGGCGCGCGAGCTCCAGCAGCTCCAGAAAGCGCGCGAGTCAAGCGCGCG 3565
 Qy 1781 GlyProThrHlsleuThrlyseThrProThrThrThrSerSerSerGluArgAlaArgAspArg 1800
 Db 3564 GGGAGCCGCGACCGACA---CCACAGATGGAGATCGGCTCGCTCGCTCCGCGCGCG 3508
 Qy 1801 AspArgGluArgAspArgAspArgGlyArgGlylyseIleleuThrSerThrThrThr 1820
 Db 3507 CCGCGGGCGTCCGACACCCCTACCGGGCGACAGAGACATCCGGGCCACCGGACAC 3448
 Qy 1821 ValGluHlsAlaPro-----IleThrArgProGlyThrGluIleSerSerGly 1836
 Db 3447 GAGTCTTCGAGCCCGCGCGCTTCACGATGTGCTGCCAGCGACGCGAGGCTGGGC 3388
 Qy 1837 SerSerGlySerSerGlyGlyGlyGlylyseSerSerArgProAlaSerHlsSerHls 1856
 Db 3387 AGTGTGACACCAAGCGTGGCGCGGAGGGGAGCCCGGAGCGCTTCACAGAGTTCGAC 3328
 Qy 1857 -AlaHlsGlnHlsSerProIleSerPro-----ArgThrGl 1868
 Db 3327 AGCTGACCGCGGTACCG---AGTCAATGCCCTGGAGGTGAAGTGGCGGTGAGCGCG 3271
 Qy 1868 MetAlaLeuGlnGlnArgProSerValleuHlsAspThrGlyMetlylyleleTh 1888
 Db 3270 ACCTGCGCGGATCGCTG----- 3252
 Qy 1888 ValAlaValGluProSerlyseProThrValleuArgSer-----ThrSe 1902
 Db 3251 -----GGCCAGAGCGCGCGCTGCTCTGCGGACCGAGCGAGCGCATCTCGGG 3199
 Qy 1902 rThrSerSerProValArgProAlaAlaThrPheProProAlaThrHlsCyProleuGl 1922
 Db 3198 GACATCGGCTCCGGTGACGCGAGT-----CCGCTCGTTCGTCGTGCTCGTGG 3148
 Qy 1922 yGlyThrLeuAspGlyValTyrrProThrLeuMetGluProValleuLeuProlyGluAl 1942
 Db 3147 GAGTTC-----CTTCGAGTCTC 3130
 Qy 1942 AspArgValAlaArgProGluArgProArgAlaAspThrGlyHlsAlaPheLeuAla 1962
 Db 3129 TCCGGGTCCGGGAGACGCTGTGCCCCACG-----CATCGGTGGCGGTCA 3082

QY 1962 sProProAlaArg-----serGlyLeu----- 1969
 Db 3081 CCGGCCCCGGCGGTGACCGGCGGCTACGGCGGCGACGAGTGAGTACGGCGT 3022
 QY 1969 ----- 1969
 Db 3021 TGGAGGGGTAGTGGGAGGTGATGCGCGCGGCTGCTCCGGTGTGGAGTGGCGTG 2962
 QY 1970 -GluProAlaSerSerProSerSerlySerGlySerGluProArgProLeuValProProValSe 1989
 Db 2961 GCGTTCACGCGCGCGGTCCGGGTGAACACCGTCGCGCGCGCGCTCGAACAACCTCCGCG 2902
 QY 1989 rGlyHisAlaThrIleAlaArgThrProAlaLysAlaMetLeuAlaProHisAlaSerPr 2009
 Db 2901 TCGGCAAGGCGCC-----GCGCGATCGCGGACCA-----CCACCAATATCATCCGCGCC 2851
 QY 2009 o-----AspProProAlaProProAlaSerAlaSerAsp----- 2020
 Db 2850 CCGAGGCACTACCGCGCATCCCGTCAGACCCCATGGAGACCACTCCACCAACGTC 2791
 QY 2021 -ProHisArgGlyLeuThrLeuSerlySerProHisSerIleGlnGluLeuGluLeuAlaGse 2040
 Db 2790 CCGACACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2732
 QY 2040 r-----LeuGlyTyrHisGlySerSerTyrSerProGlyGlyValGluProValSerPr 2058
 Db 2731 ATAGACGACCAATATCCCGGATCCCGCAACTCCCG-----GAACCCACCAACCC 2681
 QY 2058 oValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuLeuAs 2078
 Db 2680 ACCGACACGACCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2628
 QY 2078 pLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGlyProValLysLeuGly 2098
 Db 2627 ACACCTCCAC-----AACCCCGCGAAGCTCCCGCA 2598
 QY 2098 yGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGluSerSerSerPr 2118
 Db 2597 ACACCGGCTCATCAACACGGAATGAACCATGCGACACATCCACACG-----CC 2547
 QY 2118 oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnI 2138
 Db 2546 GGCACCTCCACCCATCCCGCATCGAGGACGACGACGACGACGACGACGACGACGACGACG 2499
 QY 2138 sLieserGluValIleThrGlnAspTyrThrArgHisHisPro-----GlnGlnLeuSe 2156
 Db 2498 CACCCAC-----ATCACCGGACACACGACGACGCGCGCATTCACCGCGGCAACCG 2445
 QY 2156 rAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe 2176
 Db 2444 ACACCACTCCCGCAACCTTCACCAACCCCGCGACACCGA-----CT 2400
 QY 2176 uArgArgProProSerAspLeuTyrLeuPro-----ProProAspHis----- 2190
 Db 2399 CCGAGCGCC-----CACGACGACATCCCGCCACCGGACCGGACCGCAACCA 2352
 QY 2191 ----GlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluPr 2209
 Db 2351 TCACCCACCGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2292
 QY 2209 oAsnLysThrSerValLeuGlyLysGlyLysArgGlyIleGluProValSerProProGl 2229
 Db 2291 CATACGACGACGACCTC-----CCCA 2268
 QY 2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu----- 2245
 Db 2267 CCGAATGACCA-----CACACGACACCTCCACACCGGAGCTCAAGCGCGGA 2214
 QY 2245 ----- 2245
 Db 2213 ACACGCGACCTCAACGCAACGACGAGCTGAGCAAACTCGTCGACCAACCAAC 2154

QY 2246 ----TyrArgAspGlyGluGlnThrGluProSer---ArgMetGlySerLys-Ser----- 2261
 Db 2153 CACCCAGAAACCGGACACGAAACCGACCCAGACCCGACCGGACCGGATACGCC 2094
 QY 2262 ----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAns 2280
 Db 2093 CCGACCGCGGTCCACGTGACGTGCGCG-----AACATCAATCCCGCAACG 2046
 QY 2280 eAlaMetValLysSerLysLeuGlnIleAsnLysLysLeuAsnThrHisAsnArg 2300
 Db 2045 ACCGACCATCACCTCCCGACCATGCAACAACCTCATCGACGACCTCGCAAAACCC 1986
 QY 2300 sGluProGluTyrAsnIleSerGlnProGlyThrGlu-----IlePheA 2315
 Db 1985 CACCCCGGATACAAACCGACGCGCCATACCAACCCATCGGACCGGACCGGTGAACA 1926
 QY 2315 sMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnIuH 2335
 Db 1925 ACACCCCGACCGGACGACGAC----- 1900
 QY 2335 iAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrA 2355
 Db 1900 ----- 1900
 QY 2355 sPcInTPGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA 2375
 Db 1899 --CGCCCGACCGGACACCCCGGACACCTCCCGGACGAAACCCCAAGACGACGA 1842
 QY 2375 IaSerLeuPro----- 2378
 Db 1841 CGAGCTCCCGACCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1782
 QY 2379 --AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisIleThrLeuThrSerProG 2398
 Db 1781 AACCCAGCGAATAGGACCATCCCGGATCCCGCGCCACCGCTCCACCTACCTACCGCA 1722
 QY 2398 yGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
 Db 1721 ACCGACCGGCTCGCGCG-----AACGCCCCCTCCGACCG-----GCGGACAAACA 1671
 QY 2418 IaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyA 2438
 Db 1670 GCGCGCGGAGCTCGGAAGGAGCGGCGGACGAGACGCGCGGAGTGGCGGAGGAGGCA 1611
 QY 2438 sPcysAsnArgAspThr-----ProLeuThrAsnArgValTyrGluAspA 2453
 Db 1610 GGTGTTCCAGGACACATGGCAGTTGTGCCACCGATGCCGAA-----GAACCTGA 1560
 QY 2453 rGProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
 Db 1559 CACCCCGGACGAGGATGCCCGCC----- 1534
 QY 2473 IaGlyValMetAlaSerProProProPro-GlyLeuPro---AlaGlySerGlyProLeu 2491
 Db 1533 --GGGCGAGCGGACCTCGGCGGACGCGTGCACCTTACCGGAGCTGTCCAGCG 1476
 QY 2492 AlaGlyProHisAlaTTP 2498
 Db 1475 GGATGGCCGAGAGGGGGTGG 1455
 RESULT 11
 US-08-804-198-1/c
 ; Sequence 1, Application US/08804198
 ; Patent No. 5945320
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgett, Stanley G.
 ; APPLICANT: Kuhnstoss, Stuart A.
 ; APPLICANT: Rao, Nagaraja R.
 ; APPLICANT: Richardson, Mark A.
 ; APPLICANT: Rostock, Paul R., Jr.
 ; TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:

```

1 ADDRESS:SE: PAUL R. CANFRELL, 1138
2 STREET: LILLY CORPORATE CENTER
3 CITY: INDIANAPOLIS
4 STATE: IN
5 COUNTRY: USA
6 ZIP: 46285
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: floppy disk
9 COMPUTER: Macintosh
10 OPERATING SYSTEM: Macintosh 7.0
11 SOFTWARE: Microsoft Word 5.1
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/804.198
14 FILING DATE:
15 CLASSIFICATION: 435
16 ATTORNEY/AGENT INFORMATION:
17 NAME: CANFRELL, PAUL R.
18 REGISTRATION NUMBER: 36,470
19 REFERENCE/DOCKET NUMBER: P9113
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 317-276-3885
22 INFORMATION FOR SEQ ID NO: 1:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 44377 base pairs
25 TYPE: nucleic acid
26 STRANDEDNESS: single
27 TOPOLOGY: linear
28 MOLECULE TYPE: DNA (genomic)
29 FEATURE:
30 NAME/KEY: CDS
31 LOCATION: 350..14002
32 FEATURE:
33 NAME/KEY: CDS
34 LOCATION: 14046..20036
35 FEATURE:
36 NAME/KEY: CDS
37 LOCATION: 20110..31284
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 31329..36071
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 36155..41830
44 US-08-804-198-1

```

QY 393 -----AlaValIlePro-----PrometleuTyrrAspAl 402
 Db 8108 GGTACGCAAGTTCAGAAACGGTGAAGGAGTGAACCCCGCGAGCCG-----GA 8061
 QY 402 aAaPGLInGIaRgIleLybPheIleAmeTAsnGlyLeuMetAlaAspPro-----Me 420
 Db 8060 AGGCCAGCTCCGGGTGACCCGCTCGGC-----CGAGGCTGCGCAGAGATCCCGCGCGCT 8004
 QY 420 tLveValTyrrLybAspArgIInValMetAsnMetTrpSerGIInGIInLybGIInThPh 440
 Db 8003 CCGAGCCGACAG-----CGTAGCAGAGACGACGCGCTT 7968
 QY 440 eAaRGIuLyS-----443
 Db 7967 CCGCGCGAGAGGGCCGACGAGAGCCACGCGCGCTCCGCCCTCCCTCTCCGCTCC 7908
 QY 444 -----PheMetGlnIleProLybAsnPhGIyLeuIleAlaSerPheLeuGIIn- 459
 Db 7907 CGCCCGGTGCGCTCCGCGCGCGCTCCCTGGGTGAGAGCGGACGCTGCTCCGCG 7848
 QY 460 -----ArgLybThr-ValAlaGIuCybValLeuTyrrTyrrIleuThLyS 475
 Db 7847 CATGCCCGGACGCGCGCGAAGCGCGCGAG-----7816
 QY 475 ySaEnGIaEnTyrrLybSerLeuValArgArgSerTyrrArgArgGIyLySserGIIn 495
 Db 7815 --ACGCCGTGCAAGAGAGCAGCAGGTACCGCGGTGACCGAAGCCCGACAGAGGTCTCC 7758
 QY 495 InGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIIn 509
 Db 7757 AGTCAAGTCCGCGCAGCAGAGGAGCTTCGCGCGCGCGCTGCGCTCCGCGCAGAG 7698
 QY 509 InGIInProMetProArgSerSerGIInGIInGIInLybAspGIInLybGIInLybGIIn 529
 Db 7697 TCCCGCGCGCGCCACCACTGTCGAGAGAGCTGACGCGCGCGCTCGGAGTCCAGGCG 7638
 QY 529 IaGIuLybGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIIn 537
 Db 7637 TGAAGCGCGTCCGTCGAAGACTTCGCGCGCGCTCCGAGGCAATTCGCGACCGGCC 7578
 QY 538 -----ValGIuAsnAspLybGIInAspLeuLybGIInLybThrAspAsp 553
 Db 7577 ACAGCCCGCCAGCGATGAGGCGGACGCGCTCCCTGCGCGCGCTGCGCGCGCG 7518
 QY 553 hrSerGIIn-----GluA 557
 Db 7517 CGTCGAGCGCGCGCTTGGCGCGCGCTGACGCGCTCCCTCCGTCGCCACAGCGCGA 7458
 QY 557 sPaEnAspGIInLybGIInAlaValAlaSerLybGIyArgLybThrAlaAsnSerGIInLy 577
 Db 7457 CGATGAGAGAGAGAGACAGCAGAGCGCTCGAGCGGAAGTCCGCGGTACGCTCCAG 7398
 QY 577 rGaRGIyGIyArgIleThrArgSerMetAlaEnGIInAlaAsnSerGIInLybAlaIle 597
 Db 7397 GACGG-----CGCGCGGTGA 7383
 QY 597 hrProGIInGIInSerAlaGIInLeuAlaSerMetGIInLeuAsnGIInSerSerArgTrpThr 617
 Db 7382 CTTTGTCCCGCAGTACGCGCTCCACCGCTCGAGAGTGAAGTCCAGC-----ACAC 7329
 QY 617 InGIInGIInMetGIInThAlaLybLybGIyLeuLeuGIInLybGIyArgArgTrpSerAla 637
 Db 7328 CGTCGTGACACACACCTCCGCGCTGACGTACGCGAGAGAGAGGCGCTCCCTCGGAGAG 7269
 QY 637 leAlaArgMetValGIySer-----LybThrValSerGIInCybLybAsnPhTyrr 654
 Db 7268 CCGCGAGAGAGCGCGCGAGTCCGCTCGCGAGAGTCCGAGAGCTCGAGCGC-CACT 7210
 QY 654 heAnEnTyrrLybLybArgGIInAsnLeuAspGIInIleLeuGIInGIIn-----668
 Db 7209 TCGGAACCGAGCGCGCTGACTCTCTGACGAGTTCCGCCACTCCGCGGCGCGCGCGCG 7150

QY 669 -----HisLybLeuLybMetGIInLybGIInArgAsnAlaArg-ArgLybLybLyb 685
 Db 7149 CGCGCGCCACACAG-----GAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 7099
 QY 686 AlAProAlaAlaAlaSerGIInGIInAlaAlaPheProProValValGIIn-----AspGIIn 703
 Db 7098 GCGACCTGCGCA-----CCAGGGCAGCGGTGCGCGCGCGCGCGCGCGCGCGCG 7045
 QY 704 GIuMetGIInAlaSerGIyValSerGIyAsnGIInGIInGIInMetValGIInGIInAla 723
 Db 7044 GACCACCGGTGCTCCGACAGACACAGCTCCGAGGTGTCCTCCGCGCTCTCCCGAG 6995
 QY 724 LeuHisAlaSerGIyAsnGIInValProArgGIyGIInCyS-----SerGIyProAlaThrVal 742
 Db 6984 CGCGTGGCGGGCGCGGTGAGCGCTCGAGCGGAGCGGAGCGGAGAGCGCGCGCGG- 6926
 QY 743 AsnAsnSerSerAspThrGIInSerIleProSerProHisThrGIInAlaLybAspThr 762
 Db 6925 AACCAACCGCGCGCAGTACAGCGC-----GAGAGCGCTGACCGCGAATCTGCTCCCG 6869
 QY 763 GIyGIInAsnGIyProLyS-----ProProAlaThrLeuGIyAlaAspGIy 777
 Db 6868 GCGAGCA-GCGCCCGACAGACGTACGAGACCGCGCGCGAGCACTCCGCGCGAGT 6810
 QY 778 ProProProGIyProProThrProPro-----ArgArgThrSerArgAlaProIle 794
 Db 6809 CCGCGAAGCG 6750
 QY 795 GIInProThrProAlaSerGIInAlaThr--GIyAlaProThrProPro-----810
 Db 6749 GCTGAGCGCGCGCGAGAGAGGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCG 6690
 QY 811 -----AlaProProSerProSer-----816
 Db 6689 TGATCCAGAGCGGCGCTTCGGGTACCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 6630
 QY 817 AlAProProProValValProLybGIInGIInGIInGIInGIInGIInGIInGIInGIIn 836
 Db 6629 CGCGCGCGCTCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6571
 QY 837 ValGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIInGIIn 856
 Db 6570 TCAACCGCGCGCGCGCAACAGTCCCGCGCGCTCCCGCG-----CCTGTAACGACCGCG 6517
 QY 857 LybAlaGIInGIInProValLybSerGIInCybThrGIInGIInGIInGIInGIInGIIn 876
 Db 6516 ACAGCG-----ACGCTGAGACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6463
 QY 877 GIyLybAspAlaGIInAlaGIInAlaThrAlaGIInGIyAlaLeuLybAlaGIInLyb 896
 Db 6462 GGGCGAGTCCGAGTCCGCG-----GTCTCTCGTCCGCGCGCGCGCGCGCGCGCG 6406
 QY 897 GIuGIy--GIySerGIyArgAlaThrAla-----906
 Db 6405 GCAGCGCGCGGTTGCGCGGTGACGACGAGCGCGGTGCCAAGCGAGCGGTAGCGGATC 6346
 QY 907 -----LybSerSerGIyAlaProGIInAsp 914
 Db 6345 CGCGCGCTGTCGACCTCCCGCTCCGTCAGTGGCGCGAGCGTCCGCGAGCGCGCG 6286
 QY 915 Ser-----915
 Db 6285 TCACACCGCGTACCGCGCGGTGCAAGGGGTCCAGCAGTACGCGTCCGTTGGAAGGG 6226
 QY 916 -----AspSerSerAlaThrCyS 921
 Db 6225 TAGTGAGAGGTGATGCGCGCGCGGTCTCCCGGTGAGATGCGCGGTGCGGTCCAGG 6166
 QY 922 SerAlaAspGIInValAspGIInAlaGIInGIyGIyAspLybAsnArgLeuSerProArg 941
 Db 6165 CGCGCGTCCCGGTGACACCGTCCGCGAGCGCGCGCTCGAAGCACTTCGCGTCCGAG 6106
 QY 942 ProSerLeuLeu-----ThrProThrGIyAspProArgAlaAsnAlaSerProGIIn 958

[illegible]

Db	5127	-----CCACCCGGCAGACGCCGCCGCGAGCCGAGACACCCCGGACACACCCCGGACACACCCCTCCCGGAGC	5075
Qy	1266	yHieValIleTyrgLuglybVbGlyhIbValIleuSerTyrgLuglyMetSerVa	1286
Db	5074	AAAC-----	5071
Qy	1266	lThGlnCySerlybGluAaPrgLyAArgSerSerSerdLyProProhIbGluThPAlaI	1306
Db	5070	-----CCACCAAGACGAGCAGCAGAGTCTCCACGCCACACACGAGCAGCAGCCGC	5021
Qy	1306	aProLybArgThrTyraPbMetMetGluLyArgValIGlyArgAlaIleSerSeraIase	1326
Db	5020	A-----CGTGCTCCGAAAGAGTCGCGCTCGAAGC	4991
Qy	1326	rIleGluGlyLeuMetGlyArgAlaIleProPro-----	1337
Db	4990	CAGGAATPAACSCACATCCCGGATCCGACCCACCCGCTCAGTACTCACCACCG	4931
Qy	1338	-----GluArgHisSerProhIbHieLeuLyGluGlnhIbHieIleArgGlyse	1354
Db	4930	CACGCCCTGCAGCCGCAACGCCCTCCGACCGCGGACAAACAGCCGCGAGAC	4871
Qy	1354	rIleThrgLuglyLeProArgSerTyrgValIGluAlaGlnGluAspTyrgLeuArgArgI	1374
Db	4870	AGCGCGCTCCGCGGTACCGCGGTACGAGACAGCAGCCCGCTC-GATGCTCTGACCCACT	4812
Qy	1374	uAlaIbLeuLeu-----LybArgGluG1	1382
Db	4811	CCGGCGCTCTCATGATCAGATGCGCGTTGCTCCACTGATCCGACAGCAGCAGACAG	4752
Qy	1382	yThPro-----ProProProProSerArg-----As	1392
Db	4751	CCGCCGCGCGCGCGGTCCGACCGCGGAGCCACCGCGCGCTTGCCACAGCAGGCGGA	4692
Qy	1392	pLeuThrgLualTyrgLybThrgAlaIleuGlyProLeuLybLeuLybProAlaHieG1	1412
Db	4691	CGGTCGCCGAGCAGCAGTCAGCAGCGCTGCGTGGCCGGTCCGCTGAAGTCCGCGGCA	4632
Qy	1412	uGlyLeuValAlaThrValLybGluAlaGlyArgSerIleHieGluIleProArgGluG1	1432
Db	4631	CGGTGCGGTCCGAGGAGC-----CAGCA	4605
Qy	1432	uLeuArgHieThrProGluLeuProLeuAlaProArgProLeuLybGluGlySerIleTh	1452
Db	4604	TCTTGACACAC-----CGCGGCCCGC-----GGCGGCTG-----	4571
Qy	1452	rGlnGlyThrProLeuLybTyraPbThrgAlaIleSerThThrgLySerLybHieIbAs	1472
Db	4570	-----GCTGTACCGAGCTTGAGCTTCAGCAGCAGC-----	4539
Qy	1472	pValArgSerLeuIleGlySerProGlyArgThrPheProProValHieProLeuAaPva	1492
Db	4538	GCCACAAAGGGTC-----GCCACCGCGTC-----	4514
Qy	1492	lMetAlaAaPAlaArgAlaLeuGluAlaGalaCyTyrgLuglyuSerLeuLybSerArgPr	1512
Db	4513	-----ACGCGCCAGCTGCAGCAGCGCGCTCGCTGATGATCGCCAGCGCGCTTC	4458
Qy	1512	oGlyThrAla-----Se	1516
Db	4457	CCGhACCTGGCGCTCCACCGCGCTCAGCTACCCGCGCGAGCGCGCGCTCCGAGCG	4398
Qy	1516	rSerSerGlyLySerIleAlaArgGlyAla-----Pr	1527
Db	4397	CAGCTCGGATGACGCGCTCTGCGGAGCGGTTGGAGCGGTCAAGCCGTTCGTCGAC	4338
Qy	1527	oValIleVal--ProGluLeuGlyLybProArgGlnSerProLeuThrTyrgLusPbHis	1546
Db	4337	CGTCTCGTTACCGCAGCTGCCGCGCAGCAGCCAGCAGC-----	4296
Qy	1547	GlyAlaProPheAlaGlyHieLeuProArgGlySerProValIleMetArg-----	1563
Db	4295	GGTGTCGCGCAGCGCGCGCT--CGAGACCGCTTCAGTGCCAGTACGCGAGCAGCCTT	4239

QY	1564	-----GluProThrProArgLeuGlnIuValylSerLeuSerSerLeuAlaSerGln	1561
Db	4238	CGAGCAGCAGCGTACCGTGGCGATTCGGGGAGAAAGGACATGACGAGCGCGTGGGGGCGA	4179
QY	1582	AspArgIuValLeuThrSerThrProArg-----GluIleAlaIysSerProHiserThr	1599
Db	4178	GCCGGCGGTGGCGGGAGAACTTCAGAACATCGCGGGCCCGGACATCACCGTGGCGCGCG	4119
QY	1600	ValProGluHisValProHisProIleSerPro-----	1610
Db	4118	CCGGCAGAGCC-AGCCCGCAC-----TCGGCCCGCGCGACGCGCTGCACGGCGAGGTGC	4066
QY	1611	-----TyrGluHisValLeuLeuArgIuValSerArgIuValAspLeuTyrArg	1625
Db	4065	AGCCGCCACGAGGACGAGAGACGCGCGGTGCAGCGGTAGC-----	4024
QY	1626	SerHisIleProLeuAlaPheAspProThrSerIleProArg-----GlyIleProLeu	1643
Db	4023	-----GCCGTCCGCG-----GCCGAGGGGTAGCGCATCGCGCG	3985
QY	1644	AspAlaIaAlaAlaIaTyrTyrLeuProArgHisValLeuAlaProAsnProThrTyrProHis	1663
Db	3984	GAGGCGACGCTCGG-----GTGGTCCGGGTGAGCAGCATGCCCTCG	3943
QY	1664	LeuTyrPro-----ProTyrLeuIleArgGlyTyrPro--	1674
Db	3942	---TATCGTCCGCTCCCTCGTGAAGCTCGGCGGTATCCGGTCCGGTGGCGCCACG	3886
QY	1675	AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln	1694
Db	3885	AACACTCCGGGTGGCGCGCGCGAGTACATCGGGGAGCACTCGGCGCTCCAGGGCGC	3826
QY	1695	MethHisAsnThrAlaThr-----AlaMetAlaGlnArgAla	1707
Db	3825	TCCACGACGCTTCACGACGACGAGCGCTCGCGGTCTCATGGCGGTGGCTCGGGCG	3766
QY	1708	AspMetLeuArgIuLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly	1727
Db	3765	GATATC-----CCGAAAGATCC-----GCGTGAACTCGGCGCC--	3720
QY	1728	ProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThr	1747
Db	3729	-----GAGTCGAGAAACCGCGCTCCGACATAGATCTTCGGGGCGTC	3685
QY	1748	ProGly-----ThrProAlaThrAlaMetAspArgLeuAla-----Tyr	1760
Db	3684	CCGGGCTCGGGGTGATACGCGCTTCAGAGTCCAGCGCGGTGGTGGGAGAGAGAG	3625
QY	1761	LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly	1780
Db	3624	ATGCGCTCGGGCGGAGCTCCAGCAGCTCCAGAGAGCGCGCGCATCGACGCCCGCG	3565
QY	1781	GlyProThrHisLeuThrTyrProThrThrThrSerSerSerGluValuArgAspArg	1800
Db	3564	GGAAGCGGACCGGAC-----CCACAGTAGCGATCGGGTGTGTGTGTCCGGCTCGG	3508
QY	1801	AspArgIuValArgAspArgAspArgIuValArgIuValYsSerIleLeuThrSerThrThr	1820
Db	3507	CCCCGGCGGTCCGACACCCCTACCGGGCGAGCAGAGACATCCGGGCGCCACCGCAAC	3448
QY	1821	ValGluHisAlaPro-----IleThrArgProGlyThrGlnHisSerGly	1836
Db	3447	GAGTCTTCGGCGCGCGCGCTTCACAGATGTCTCGCGACGACGCGGGGTGGC	3388
QY	1837	SerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis	1856
Db	3387	AGCTCGTAGACACAGCGTGGCGGCGAGGGGAGACCCCGAGGCGCTTCAGAGGTCCGAC	3328
QY	1857	-AlaHisGlnHisSerProIleSerPro-----ArgThrGln	1868
Db	3327	AGCTCGACCGGGGTACCG--AGTCAATGCTCGGAGGTGAAGTGGGTCCAGGCGG	3271

QY	1668	nAeRAlaLeuGlnGlnAaGPrProSerValLeuHiaAanthrGlyMetuYvGlyLeIeth	1868
Db	3270	ACCTCGCGGGAGTGGCT	3252
QY	1888	rAlaValGluProSerLeuProThiValLeuArgSer	1902
Db	3251	-----GCGCCAGACCGCGCGCGTCTCGCGAGACCGAGCGGACTCTCGGGG	3199
QY	1902	rThiSerSerProValArgProAlaValThrPheProProAlaThiHisCysProLeuG1	1922
Db	3198	GACATCGGCTCGCGGAGACCGCGAGT-----CGGCTCGTGGTCTCGTCTCGCTGG	3148
QY	1922	yG1rThiLeuAerGlyValrProThrLeuMetGluProValLeuLeuProLyvGluAl	1942
Db	3147	GAGCTCC-----CTCGGTCTC	3130
QY	1942	aProArgValAlaArgProGluAArgProArgAlaAerThrGlyHisValPheLeuAlaLy	1962
Db	3129	TCCGGGTCGGGGAGACGCTGTGCTCCACCG-----CATCGTGGCGGTACA	3082
QY	1962	sProProAlaArg-----SerGlyLeu-----	1965
Db	3081	CCCGCGCGCGGTGCGCGGTGACCGCGCGGTACCGCGCGGTACCGAGTGGTACCGGCT	3022
QY	1969	-----	1965
Db	3021	TGGAAGGGGTAGTGGAGGTGCGATGCGCGCGGTCTCCGCTGTGGATGCTCGT	2962
QY	1970	-GluProAlaSerSerProSerLyvGlySerGluProArgProLeuValProProValSe	1989
Db	2961	GGCTGCAGCGCGCGGTGCGGTGACACGCTGCGCGAGCGCGCTGGACACTCCCGC	2902
QY	1989	rGlyHisAlaThrIleAlaArgThrProAlaLyvAsnLeuAlaProHisAlaSerPr	2009
Db	2901	TCCGCAACGGCGCGCG-----GCCGATGCGCGGACCA-----CCACACTATCAACGGCC	2851
QY	2009	o-----AspProProAlaProProAlaSerAlaSerAp-----	2020
Db	2850	CCCAAGCACTACCGCGCATCCCGTACGACCCCAAGCAAGCACTCCACCAACAGTCT	2791
QY	2021	-ProHisArgGlyvThrGlnSerLyvProPheSerIleGlnLeuLeuGlyLeuAArgSe	2040
Db	2790	CCCAACCCCAACCAACGACGACACCCCAACCCCA-----TCCGCGAAAGCAACCGCTCCCGC	2732
QY	2040	r-----LeuGlyThrHisGlySerSerTyserProGlyGlyAlaGluProValSerPr	2058
Db	2731	ATGACGACACCAATACCCCGGATCCCAACTCCCC-----GACCCACACACCC	2681
QY	2058	oValSerSerProSerLeuThrHisAerLyvGlyLeuProLyvHisLeuGlnGlyLeuA	2078
Db	2680	ACCGGACACACCGGACACCAACCAACCGGCG-----CACCGACCGAATCCA	2628
QY	2078	pLyvSerHisLeuGlnGlyLeuAArgProLyvGlnProGlyProValLyvLeuGlyG1	2098
Db	2627	ACGACTCCAC-----AACCCCGCGAATCCCA	2598
QY	2098	yGluAlaAlaHisLeuProHisLeuAArgProLeuProGluSerGlnProSerSerSerPr	2118
Db	2597	ACACCGGCTCATACACCGAAGAAACCATGCGACACTCCACCG-----CC	2547
QY	2118	oLeuLeuGlnThrAlaProGlyValLyvGlyHisGlnAArgValAlaThiLeuAlaGlnI	2138
Db	2546	GGCACTCCACCCCATCCCGCATCGAGGAGAACACCACTC-----CAGA	2499
QY	2138	gHisSerGluValIleThrGlnAerTyrrThrArgHisHisPro-----GlnGlnLeuSe	2156
Db	2498	CACCCAC-----ATCACCGGACAAACCAACCGACCGCGCGCATTCACCGCGCAACCG	2445
QY	2156	rAlaProLeuProAlaProLeuTyserPheProGlyAlaSerCysProValLeuAerLe	2176
Db	2444	ACACCCACTCCCAACCCCTCAACAAACCCCGGACACCGA-----CT	2400
QY	2176	uArgArgProProSerAerLeuTyrrLeuPro--ProProAerHis-----	2190

Db	2399	CCGAGGCCCC-----	:::	-----CACCGACCAATCCCCCACCACCGGGAACCCACCA	2352
Qy	2191	----GlyAlaProAlaArgIyserProHisSerArgSerAlaValTyrProLeuLeu		-----GlyAlaProAlaArgIyserProHisSerArgSerAlaValTyrProLeuLeu	2209
Db	2351	TCAACCCACCCCGCGGCACACCAACCGCACCGCATCCACCAACGACAAACCCCGGCA		-----TCAACCCACCCCGCGGCACACCAACCGCACCGCATCCACCAACGACAAACCCCGGCA	2292
Qy	2209	oAsnIysThrSerValLeuGlyGlyGlyGlyLeuAbpGlyIleGluProValSerProProGly		-----oAsnIysThrSerValLeuGlyGlyGlyGlyLeuAbpGlyIleGluProValSerProProGly	2229
Db	2291	CATACGGACGACCACTCC		-----CATACGGACGACCACTCC	2268
Qy	2229	uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu		-----uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu	2245
Db	2267	CCGAATGACCCAA-----		-----CCGAATGACCCAA-----	2214
Qy	2245	-----		-----	2245
Db	2213	ACAACGCCACTCCAAAGCAAAACAAGAGGCTGAGAAATCCGTCCGACCAACCAAC		-----ACAACGCCACTCCAAAGCAAAACAAGAGGCTGAGAAATCCGTCCGACCAACCAAC	2154
Qy	2246	----TyrArgAbpGlyGlyGlnThrGluProSer---ArgMetGlySerLys-Ser---		-----TyrArgAbpGlyGlyGlnThrGluProSer---ArgMetGlySerLys-Ser---	2261
Db	2153	CACCCACGAACCGGAMACGAACCGACCGACACCGGACCCGACACCCCATCAGGCC		-----CACCCACGAACCGGAMACGAACCGACCGACACCGGACCCGACACCCCATCAGGCC	2094
Qy	2262	-----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn		-----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn	2280
Db	2093	CGGACACCCGCTCCACGTCGACGTCCGCG-----		-----CGGACACCCGCTCCACGTCGACGTCCGCG-----	2046
Qy	2280	eAlaMetValIysSerLysGlyGlnIleAsnLysLysLeuAsnThrHisAsnArgA		-----eAlaMetValIysSerLysGlyGlnIleAsnLysLysLeuAsnThrHisAsnArgA	2300
Db	2045	ACCGACCATTCACCTCCCCCACCACATTCGACAAACCTCATTCAGACACCTCCGCAAAACACC		-----ACCGACCATTCACCTCCCCCACCACATTCGACAAACCTCATTCAGACACCTCCGCAAAACACC	1986
Qy	2300	sngIuProGluTyrAsnIleSerGlnProGlyThrGlu-----		-----sngIuProGluTyrAsnIleSerGlnProGlyThrGlu-----	2315
Db	1985	CACCCCGCGCATACCAACCCAGCCCATACCAACCCACTGGCGCACCTGACCCGTGAACA		-----CACCCCGCGCATACCAACCCAGCCCATACCAACCCACTGGCGCACCTGACCCGTGAACA	1926
Qy	2315	smeProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluH		-----smeProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluH	2335
Db	1925	ACACCCCAACCCACCCGACACGCC-----		-----ACACCCCAACCCACCCGACACGCC-----	1900
Qy	2335	IsAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrA		-----IsAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrA	2355
Db	1900	-----		-----	1900
Qy	2355	spgIuTyrGluIserProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA		-----spgIuTyrGluIserProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA	2375
Db	1899	--CGCGCGACCGCACACCCCGACACCTCCCGCGACGAAACCCACCAAGACAGCGGA		-----CGCGCGACCGCACACCCCGACACCTCCCGCGACGAAACCCACCAAGACAGCGGA	1842
Qy	2375	IsaSerLeuPro-----		-----IsaSerLeuPro-----	2378
Db	1841	CGAGTCCCCACGCCACACACCGCACACCGCACCGGTCTCGAAGAGTCCGCGCTCG		-----CGAGTCCCCACGCCACACACCGCACACCGGTCTCGAAGAGTCCGCGCTCG	1782
Qy	2379	--AlaAlaMetProIleThrAlaAlaAbpGlyArgSerAspHisThrLeuThrSerProG		-----AlaAlaMetProIleThrAlaAlaAbpGlyArgSerAspHisThrLeuThrSerProG	2388
Db	1781	AAGCCAGGAATAGCCACATCCCGCGGATCCGCGCCACCCCGCTCCACGATCATCCCA		-----AAGCCAGGAATAGCCACATCCCGCGGATCCGCGCCACCCCGCTCCACGATCATCCCA	1722
Qy	2398	IyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA		-----IyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA	2418
Db	1721	ACCGACCGCGTGGCGCGC-----		-----ACCGACCGCGTGGCGCGC-----	1671
Qy	2418	IaProGlyLeuAlaSerGlyAspArgProProSerValHisSerGlyGlyA		-----IaProGlyLeuAlaSerGlyAspArgProProSerValHisSerGlyGlyA	2438
Db	1670	GCGGCGGAGCGTCCGGAAGCGAGGCGGCGACGAGAACCGCGGCGTGGCGCGGAGGCGA		-----GCGGCGGAGCGTCCGGAAGCGAGGCGGCGACGAGAACCGCGGCGTGGCGCGGAGGCGA	1611
Qy	2438	spCyAsnArgArgThr-----		-----spCyAsnArgArgThr-----	2453
Db	1610	GGTGTTCAGAGCACCATGGCATGTTGGTCCACCGATCCGGAAC-----		-----GGTGTTCAGAGCACCATGGCATGTTGGTCCACCGATCCGGAAC-----	1566
Qy	2453	rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA		-----rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA	2473
Db	1559	CACCCCGACGACGGGTGCGCGGCC-----		-----CACCCCGACGACGGGTGCGCGGCC-----	1534
Qy	2473	IaGlyValMetAlaSerProProProPro-GlyLeuPro--AlaGlySerGlyProLeu		-----IaGlyValMetAlaSerProProProPro-GlyLeuPro--AlaGlySerGlyProLeu	2491

Db		1333	-GGGCGAGCGCAGCTCCGGCGGCACCGGTCTGCACTTCACCGCGGAGCTGGTCCAGCG	1476
QY		2492	AlaGlyProHisIAsLaTPR	2498
Db		1475	GGATGCCCGAGAGCGGGGTGG	1455
 RESULT 12 US-09-103-840A-2				
	/ Sequence 2,	Application US/09103840A		
	/ Patent No.	6294328		
	/ GENERAL INFORMATION:			
	/ APPLICANT:	FLEISCHMAN, Robert D.		
	/ APPLICANT:	WHITE, Owen R.		
	/ APPLICANT:	FRASER, Claire M.		
	/ APPLICANT:	VENTER, John C.		
	/ TITLE OF INVENTION:	DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM		
	/ FILE REFERENCE:	24366-20007.00		
	/ CURRENT APPLICATION NUMBER:	US/09/103,840A		
	/ NUMBER OF SEQ ID NOS:	2		
	/ SOFTWARE:	Patentin Ver. 2.1		
	/ SEQ ID NO 2			
	/ LENGTH:	4403765		
	/ TYPE:	DNA		
	/ ORGANISM:	Mycobacterium tuberculosis		
	/ FEATURE:			
	/ OTHER INFORMATION:	CDC 1551		
	/ OTHER INFORMATION:	"n" bases at various positions throughout the sequence		
	/ OTHER INFORMATION:	represent a, t, c or g		
	US-09-103-840A-2			
 Alignment Scores:				
	Pred. No.:	8.85e-19	Length:	4403765
	Score:	668.50	Matches:	617
	Percent Similarity:	32.38%	Conservative:	255
	Best Local Similarity:	22.91%	Mismatches:	1044
	Query Match:	5.06%	Indels:	793
	Gaps:	3	Gaps:	113
 DB: US-09-522-753-5 (1-2517) x US-09-103-840A-2 (1-4403765)				
QY		4	SerThrGlnLeuValAlaGlnThrTrpArgAlaThrLurProArgTyPProProHisSer	23
Db		2794906	TGACGGAAAGCCCGCATGCTGCTGCGCGGCGCACCCGAATTACCGCATTCACCGCTGACG	2794965
QY		24	-----LeuSerTyPProValGlnIleAlaGThrHisThrAsp	36
Db		2794966	AGCGGGAACCGSGTATGCAAGCTGTCACACATTTCCGTCACCGCGCACGCGACC---	2795022
QY		37	ValGlyLeuLeuGluTyGlnHisHisSerArgAAspTyAlaSerHisLeuSerProGly	56
Db		2795023	-----CGCGACGCTCAAGTTGATCGCCACGCGGA	2795052
QY		57	SerIleIleGlnProGlnArgArg-ArgProSerLeuLeuSerGluPheGlnProGlyAs	76
Db		2795053	-----ATGCCGTGAGTCCGCTGCGCAATCTCG--TCACCAAGGCG	2795091
QY		76	nGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyPLeuProGlyLeuGl	96
Db		2795092	AGTTGTCGCGCATATCTCGAGTTCCGGGCGCGCCGCTTACAAGCGCGGTGGTA----	2795146
QY		96	YLysSerGluMetGluPheIleGluSerLySargProArgGluGluLeuLeuProAspPr	116
Db		2795147	-----CAACTGATCGCTCGCGCGTGCC	2795169
QY		116	cLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuth	136
Db		2795170	CCAGCGGGGGAACCCGCGCAATCTGCTCACCGCGCAC--	2795206
QY		136	rLYAspArgSerIleuThrGlyLySLeuGluProValSerProProSerProProHisIsth	156
Db		2795207	-CGGATCGGTCCCCGGCTGGTCCC-----	2795230

QY 156 rAspProGluLeuGluLeuValProProArgLeuSerLeuGluLeuLeuGlnAme 176
 Db 2795231 -----CAATACCTTCAGCGCTGGGGCAACGCCCGAG----- 2795260
 QY 176 tAspArgValAspArgGluLeuIlethMetValGluGlnGlnIleSerLeuLeuValy 196
 Db 2795261 -CAACCGAGCATACAGCCCGCACCGGTCGAGCAAGTCCTCCAGTTGGCAGCACTA 2795319
 QY 196 sGlnGlnGlnLeuGluGluGluValAlaValProProGluProGluValProValSerPr 216
 Db 2795320 CCAAG-----CATGGCGCGGTGCGCA 2795340
 QY 216 oProProIleGluSerLeuSerHisArgSerLeuValGlnIleIleTyrAspGluAmaArgLy 236
 Db 2795341 TAGCCGCAACAAATGTCGTCCACCGTCAGAGGGCGCGGTATCCGGCAACCCCAAAACC 2795400
 QY 236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro-- 255
 Db 2795401 GCGCGCGCGCATCGACACGCGCGGTGTCGATCGCGCGCACGGTTGCATACCAAA 2795460
 QY 256 -----LeuTyrAs 258
 Db 2795461 CCCGTCGGAATACCTCGCAACGCGCTCGGACCTGTGTCCAGCGCGTCTTTC 2795520
 QY 258 nGlnProSerAspThrArgGlnTyrHisGluAmaIleLysIleAsnGlnAlaMetArgLy 278
 Db 2795521 CGACCCCGCGCACCGGTAAAGGTGACCCACGTTGAGCTCCAGACGCCACGAC-- 2795578
 QY 278 sLysLeuIleLeuTyrPheLysArgArgAmaHisAlaArgLysGlnTyrLysGlnLysPhe 298
 Db 2795579 -----TTGCGGCACCTTCCTCGACGCGCCACCGCCACCGACAGTGA----- 2795617
 QY 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
 Db 2795618 -----CTGGCGCGAAGACAGTGGCGACCAACGA 2795646
 QY 318 naenProArgArgAlaValLeuGluSerLysValArgGluTyrTyrGluLysGlnPhePr 338
 Db 2795647 CTTTCGGGGTCCGACGCGGGGGAACCGCTTGTGCATACAGGTGACACAGCTGCACCA 2795706
 QY 338 oGluIleArgLysGlnArgGluLeuGln-----GluArgMetGlnSerArgVa 354
 Db 2795707 CCCGTTCCGGTCGCGGACGGTCTCCAGCGGTAGTACCGAGAGTGTTCAGCACCGCT 2795766
 QY 354 l---GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlnIleGluLse 373
 Db 2795767 CCTTGGGCGAG-----CAGGTGAG 2795784
 QY 373 rGluIleIleAspGlyLeuSerGluGlnGlnIleuAsnLeuGluLysGlnMetArgGlnLeuAl 393
 Db 2795785 CAACGAGATCGCTGTAGTTCGCCGACAAACGAGTCTG---GCCCGCTGGGCCAGCTCGC 2795841
 QY 393 aValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIle----- 410
 Db 2795842 GCAG---CCGGCGGTGCGGTGCATGTCGCGCTACGACAGTTGCTCTGCGGTGACG 2795898
 QY 411 -AsnMetAsnGlyLeuMetAlaAsp-----ProMetLy 421
 Db 2795899 ACACCTCCCGGTGTGATGCCATGCGACGCGGATCGGTGCGACGGCGGCCGCTGCA 2795958
 QY 421 sValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGlnThrPheAr 441
 Db 2795959 AGCCCAAGGGCGACGCCACGCGCTCGATCCGGGGGGAACGCCACCAAAAGCTGTCCG 2796018
 QY 441 gGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLy 461
 Db 2796019 CTTGCGCTGTTTCAGACCGGCAAAACCCGCGGTCTCGCAACCAATTCCGTGACGGTTC 2796078
 QY 461 sThrValAlaGluCysValLeuTyrTyrTyrLysThrLysLysAsnGluAsnLysLysSe 481
 Db 2796079 G----- 2796079

QY 481 rLeuValArgSerTyrArgArgArgLyLysSerGlnGlnGlnGlnGlnGln 501
 Db 2796080 -----GTCCAGTTTGGCGATCGCGCTGTCGTTCAGCTGACACCCCGGAGGGTCCGCC 2796135
 QY 501 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 521
 Db 2796136 CTTGATATTCGCGACGACGACAGCTCA-CCGTGCCG----- 2796171
 QY 521 pGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 541
 Db 2796172 -----TCGTCACAGCTGCTTACACCATTCGCGCTCCAGTCCACAGTACCACT-- 2796222
 QY 541 pLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAspGluLy 561
 Db 2796222 ----- 2796222
 QY 561 sGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyAr 581
 Db 2796223 -----CGACGCGCGGTGA 2796236
 QY 581 gIleThrArgSerMetAlaAsnGlnAlaAsnSerGluAlaIleThrProGlnGlnSe 601
 Db 2796237 -ATCTTGCTCATGCTAGCC-----AGCATCGAGCCAGCGCTAG 2796274
 QY 601 rAlaGluLeuAlaSerMetGluLeuAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 620
 Db 2796275 CGCATTAACATCGGACACCTGCGCTTGAATCTCTGGAATCTCTTGGCCGCGGTTCATGTG 2796334
 QY 621 -----GluThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTrpSe 635
 Db 2796335 TTCTACGCGCTTATGTCATGATGATTCGCGACCGCGGTGTCATTCGCGCGCGG-ACCTACAG 2796393
 QY 635 rAlaIleAlaArgMetValGlySerLysThr-----ValSerGlnCysLysAsnLysPhe 654
 Db 2796394 T-----GCCGCTGCTCCGCTGCGCAATTGATGCCAGTGC----- 2796433
 QY 654 eAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
 Db 2796434 -----TGGCACATTCGCCGCAAGAAC 2796456
 QY 674 uLysGluArgAsnAlaArgArgLysLysValAlaProAlaAlaAlaSerGluGluAl 694
 Db 2796457 GACACGACCGCATCGTGGGGGTG-CCGTGCAAGTGGCCGCTGACCGATCGCGGAGTGC 2796515
 QY 694 a-----AlaPheProProValValGluAspGluGluMetGluLysSerGlyVa 710
 Db 2796516 ACCGCGCCGATGGGGTTCGAAAGCTAGGAAATGATGAGAGGGCGCGCTCGGC 2796575
 QY 710 l-----SerGlyAsnGluGluGluMetValGluGluAlaG 722
 Db 2796576 CGCCTTCCGCGTAACCCCAACATCGGAAACGAAACGAGCTAGCGTCCCGCGCGCGCAG 2796635
 QY 722 uAlaLeuHisAlaSerGly--AsnGluValProArg-----GlyG 725
 Db 2796636 AGCGACCATTCGCGGTGGCGCCAAACAGACGATCCGATCTCTGACCATGAGTGG 2796695
 QY 735 uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGlnGlnIleProSerProH 755
 Db 2796696 GTGTTCTGCTACGTAGCTTA-----ACGGAGATGAGAGGCGCGCTTCGCGC- 2796744
 QY 755 sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGluAl 775
 Db 2796745 -----CTTACACGCGC----- 2796756
 QY 775 aAspGlyProProProGlyProProProThrProProArgArgThrSerArgAlaProIleG 795
 Db 2796757 -GAACCGCGGAGCCAGGTGCGCTCCCTCGTTCGCGAGCACCCGACTCGCGCACGA 2796815
 QY 795 uProThrProAlaSerGluAlaThrGlyLysProProPro-----ProProAlaProPr 813
 Db 2796816 GCGGACGCGCGGAGAGTCAAGCACCGGACCGGCTCGGCTCACTTCGCGCGCGC 2796875
 QY 813 oSerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaAl 833

QY 1509 sserArgProGlyThrAlaSerSerGlySerIleAlaArgGlyAlaProVal11 1529
 Db 2798632 -GGTCCCGCCCGAACCCCGTTGCTCCCGGTGGCCGAGCGGAGCGATC----- 2798682
 QY 1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyrgLysPheHisGlyAlaPr 1549
 Db 2798683 -----GAACCAAGCGCGCCGCAATTGGCCCGGTGGCGCGCGCCGATCGCCAC 2798735
 QY 1549 orPheAlaGlyHisLeuProArgGlySerProValThrmethArg-----GluProTh 1566
 Db 2798736 A-----TTGCCCGCGCGCCGACCATCGCCGACCGCGCGCGATTCGCGCGTC 2798783
 QY 1566 rProArgLeuGlnGlnGlySerLeuSerSerSerGlyAlaSerGlnAparGlyLysLeuTh 1586
 Db 2798784 GCGTCCGCTGCC-----CC 2798798
 QY 1586 rSerThrProArgGlyLysLeuAlaLysSerProHisSerThrValProGlnHisPheProH 1606
 Db 2798799 GCGCGCCCGCGCGACGCTGGCTTCGCGACCGATTCACCGATTCGCGCGCGCT 2798858
 QY 1606 sProIleSerProTyrgLysLeuLeuAlaArgGlyValSerGlyValAparLeuTyrgSe 1626
 Db 2798859 GCGCG-----GCGCCGCTGG-----TCGCTGGCGCTCCACACCGCTG---GCCGC 2798897
 QY 1626 rHisIleProLeuAlaPheAparProThrSerIlePro-ArgGlyIleProLeuAparAla 1646
 Db 2798898 GCGGCTCCGCTTCGCGCTGTCGCGCGCCAG-----CCGCGAGCGCTCCGCGCGCGGCTC 2798954
 QY 1646 AlaAlaAlaTyrgTyrgLeuProArgHisLeuAlaProAparProThrTyrgProHisLeuTyrgP 1666
 Db 2798955 ACCGCGCTACACCGCGCTGCGCGCATCA-CTGCATGACCGCGCTGACCGCGCTGGCGC 2799013
 QY 1666 ro-----ProTyrgLeuIleArgGlyTyrgProAparThrAlaLeuGlyApar 1682
 Db 2799014 CGGCTCGCGCGAGCGCGCG-CTCACCGCCACACCGCGCGCGCGCTGCT----- 2799062
 QY 1682 rGlnThrIleIleAparTyrgIleThrSerGlnGlnMetHisAparThrAlaThra 1702
 Db 2799062 ----- 2799062
 QY 1702 LeuAlaGlnAlaArgAlaAparMetLeuAlaGlyLeuSerProArgGlnSerSerLeuAla 1722
 Db 2799063 -----GCGGACCTGACCGCGGCTGCGCGCGCGCGCGCGCATC----- 2799104
 QY 1722 euAparTyrgAlaAlaGlyProArgGlyIleIleAparLeuSerGlnValProHisLeuProV 1742
 Db 2799105 -----ACCGCGCGCGCG----- 2799116
 QY 1742 AlLeuVal-ProProThrProGlyThrProAlaThrAlaMetAparGlyLeuAlaTyrgLeu 1761
 Db 2799117 --GCGTACCGCGCGACCGCGCGCGCGCGCGCGCTGCGCG----- 2799153
 QY 1762 ProThrAlaProGlnProPheSerSerSerArgHisSerSerSerProLeuSerProGlyGly 1781
 Db 2799154 CTTTCCCGCGCGGCGCTCCACAGCGCTGCGCGCTGCGCGCGCGCGCGCGCG----- 2799207
 QY 1782 ProThrHisLeuThrLysProThrThrThrThrSerSerSerGlyAlaGlyAparGlyApar 1801
 Db 2799208 -----TTGCCCGCGCGCGCG-ACCACThAGCTCCGCG----- 2799236
 QY 1802 ArgGlyAparGlyAparGlyAparGlyAlaGlyLysSerIleLeuThrSerThrThrThrVal 1821
 Db 2799237 -----ACGACCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGCGCGGCTCCACCGCTGCC 2799284
 QY 1822 GlnHisAlaProIleThrPargProGlyThrGlnGlnSerSerGlySerSerGlySerSer 1841
 Db 2799285 CGACACCGACCG-----ACCGCGCGCGCGCGCGCGCTCGCG 2799317
 QY 1842 GlyGlyGlyGlyGlySerSerSerArgProAlaSerHis----- 1854
 Db 2799318 -----GCGACCGCGCGCGCGCGCG-CCGCGCTGCCGATCGCTCCGACCGCGCGCTTCGCGC 2799370

QY 1855 -----SerHisAlaHisGlnHisSerProIleSerProArg----- 1866
 Db 2799371 CACTACCTAACCGCGAAGCTCCCGCGACACCGGTACCGCGCGCGCGCGCGCGCGCGCGAC 2799430
 QY 1867 ThrGlnAparAlaLeuGlnGlnArg--ProSerValLeuHisAparThrGlyMetGlyGly 1885
 Db 2799431 CGCGACCGCGCGCTTCGACCGCGCGCTTCGCTTCGCTTCGCGCGCGCGCGCGCGCGCGCG 2799490
 QY 1886 IleIleThrAlaValGluProSerLysProThrValLeuAparSerThrSerSer 1905
 Db 2799491 TGCCG-TGGGCG-----CCGCTTCACCTACGCTTC-----GCG 2799522
 QY 1906 ProValArgProAlaAlaIleThrPheProProAlaThrHisGlySerProLeuGlyGlyThrgLeu 1925
 Db 2799523 CCAATTCGACCG 2799582
 QY 1926 AparGlyValTyrgProThrLeuMetGluProValLeuLeuPro-----LysGlnAla 1942
 Db 2799583 CCGCGCGTACCG 2799642
 QY 1943 ProArgValAlaArgProGlu-----ArgProArgAlaAparThr 1955
 Db 2799643 CCGTTACCGCGCGCTTCG 2799702
 QY 1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPro 1975
 Db 2799703 ACACCGCTGGCG 2799747
 QY 1976 SerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla 1995
 Db 2799748 GCGCGTGGGTACACCATCATCG 2799801
 QY 1996 ArgThrProAlaLysAparLeuAlaProHisAlaSerProAparProAlaProPro 2015
 Db 2799802 ATGTGCG 2799855
 QY 2016 AlaSerAlaSerAparProHisArgGlyLysThrGlnSerLysProPheSerIleGlnGlu 2035
 Db 2799856 GACACACCGCTCCCGCGCA-----ACACCTGCG----- 2799882
 QY 2036 LeuGlyLeuArgSerLeuGlyTyrgHisGlySerSerTyrgSerProGlnGlyValGluPro 2055
 Db 2799883 CTGCGCGCGCGCA-----CCGCGAGGTGCGCTCACCG 2799912
 QY 2056 ValSerProValSerSerProSerLeuThrHisAparLysLeuTyrgLeuProHisLeuGlu 2075
 Db 2799913 CAGCGCCG 2799966
 QY 2076 GlnLeuAparLysSerHisLeuGlnGlyGlyLeuAparProLysGlnProGlyProValLys 2095
 Db 2799967 CCGTTCCCGCGCGCGCATATGACGACGACGATTAACCGGTGACCTCGGTACCA----- 2800020
 QY 2096 LeuGlyGlyGlyAlaHisLeuProHisLeuAparProLeuProGlnSerGlnProSer 2115
 Db 2800021 -----CCG 2800071
 QY 2116 -----SerSerProLeuLeuGlnThrAla 2123
 Db 2800072 CCGCGCGTGGCG 2800131
 QY 2124 ProGlyValLysGlyHisGlnArgValAlaThrLeu-----AlaGlnHisIleSerGlu 2141
 Db 2800132 CCGGCGGCTACCGATGAACCAAGAGATGCGCTCATATCGGCTATGCGCGCGCGCGCGCGCG 2800191
 QY 2142 ValIleThrGlnAparTyrgThrArgHisAparProGlnGlnLeuSerAlaProLeuPro-- 2160
 Db 2800192 GCTCCGCTACCGCGCATGACG 2800251
 QY 2161 --AlaProLeuTyrgSerPheProGlyAlaSer-----GyProValLeu 2174
 Db 2800252 GACCGCGCGTTCACCGCATGACG 2800311
 QY 2175 AparLeuAparGlyProProSerAparLeuTyrgLeuPro-ProProAparHisGlyAlaProAl 2194

Db	2800312	CCGCCGTTGCCGCCGCTACACCGGCGCCGTCGTTGCCGCCGCCGACACCATTCG- 	2800370
Qy	2194	aAaGlySerProHisSerGluGlyGlyValysArgSerProGluProAsnIysThrSera 	2214
Db	2800371	GCGAGCACCGGCGCGCCGACGCGCCCGCGTCCGCGACGCCACCCAGG----- 	2800422
Qy	2214	lLeuGlyGlyGlyAlaMetGlylLeuProValSerProGluGluMetThrGluPro 	2234
Db	2800423	-----CCGGGTCGCGACCGTGGACCGCGCGCGCGCC 	2800454
Qy	2234	oGlyHisSerArgSeraIaValTyProLeuLeuTyArgAspGlyGluGlnThrGluPro 	2254
Db	2800455	GTCACCGCGCGCGCCCACTGCCACGCTG-----CCGCGGACCGCC 	2800499
Qy	2254	oSerArgMetGlySerIysSerProGlyAsnThrSerGlnProProAlaPhePheSerIy 	2274
Db	2800500	GTCACACAGGCGCTCCACCCAGCCGCCACCCCGCGACGCGCGCGTCTCGCGCGGCG 	2800559
Qy	2274	sLeuThrGluSerAsnSeraIa-MetValIysSerIysGlnGluIleAsnIysIysL 	2294
Db	2800560	GACATACCCCATATAGTCCAGCCGACCCCGCGACACCCCGCGCGCGCCACCGCGCGTT 	2800619
Qy	2294	euaSerThrHis-----AsnArgAsnGluProGluTyArgAsnI 	2306
Db	2800620	TCCGCGCGACCGCCATTGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 	2800678
Qy	2306	lSerGlnProGlyIThrGluIlePhe-AsnMetProAlaIleThrGlyIThrGlyLeuMet 	2325
Db	2800679	-----CCGCGCCATTATTTCGGTAGCCCGGACCGCGCGCGCGCGCGCGCGCGCG 	2800727
Qy	2326	ThrTyArgSerGlnAlaValGlnGluIleAsnIleThrAsnMetGlyLeuGluAlaIle 	2345
Db	2800728	ACCGGACACCG 	2800781
Qy	2346	lLeargIysAlaLeuMetGlyIysThrArgAspGlnThrGluGluSerProProLeuSera 	2365
Db	2800782	-----GAGAGCCACCTTCCCGACACCC 	2800805
Qy	2366	AsnAlaPheAsnProLeu---AsnAlaSeraIaSerIleProAlaIleMetProIleThr 	2384
Db	2800806	GTTGGCGGCCACCCACCGCGGACCGCGCATCGCCCCCGTCCGACCTCGCAACCGTCA 	2800865
Qy	2385	AlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyIysAlaIys 	2404
Db	2800866	GCGCGC-----ACCGTCGTCGACCGCGGCAACACACGCGCGCGCATC 	2800907
Qy	2405	ValSerGlyArgProSerSerArgIysAlaIysSerProAlaProGlyLeuAlaSerGly 	2424
Db	2800908	CTGACCGGCGCGTAGACCGTGGCGCGCGGACCATACACCCACCGCGCGCGCGCGC 	2800967
Qy	2425	-----AspArgProSerSerValSerSeraIaHisSerGluGlyAspCysAsnArgArg 	2442
Db	2800968	CCCATGACCAACACACCCCGGTTACCGC---CGCGCGCGCGCGGACACACACGAGCC 	2801024
Qy	2443	ThrProLeuThrAsnArgVal-TyrGluAspArgProSerSeraIaGlySerThrProPh 	2462
Db	2801025	ACCGGACACACCGGCGCGCGTTCGCAACAGCGCGCGTGGCCACCGCGCGCGCGGCG 	2801084
Qy	2462	ePro-----TyrAsnProLeuIleI 	2468
Db	2801085	GCCACACACGCGCCACCCATACCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGT 	2801144
Qy	2468	eMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySe 	2488
Db	2801145	CCCGCGCGGTACTCCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATCC 	2801204
Qy	2488	rgIysProLeuAlaGlyPro 2494 	
Db	2801205	TAAACCGCGCGCGCGCGCG 2801223 	

```

Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO. 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Alignment Scores:	
Pred. No.:	1,28e-18
Score:	665.00
Percent Similarity:	33.04%
Best Local Similarity:	22.44%
Query Match:	5.03%
DB:	3
US-09-522-753-5 (1-2517) x US-09-103-840A-1 (1-4411529)	
	Length: 4411529
	Matches: 627
	Conservative: 296
	Mismatches: 111
	Indels: 776
	Gaps: 122

[illegible]

QY	775	1aapBglyProProRobglyProProthirProProAaGATgThSeraArglaAProIleg	795
Db	2801304	--GAACCCCGGAGGCAAGGATGCGCCCTCCGCTGCGGACCCACCGCATCCGCGACG	2801361
QY	795	1uProthProAlaSerArglaAaThrArglaAProthPro-----ProProAlaProp	813
Db	2801362	AGCCCAAGCCGCGGAGGTCAAGCACCGGACCGCGTCCGCTCACTTCGCGCCGC	2801421
QY	813	roSerProSerAlaProProProValProLySgluGluSgluGluThraIaA	833
Db	2801422	CGTCCCGCGGTACACGCGCGCGAGTGTGCG-AGCGGAGGAGGCGCGGACCGCGGAG	2801480
QY	833	1aAlaProPro-----ValGluGluGluGluGluGluGluProProAlaAag	849
Db	2801481	CGCGCGGTCCCGCGCTCCGCTCCGTCAACACCGGGGGGCCACCGTCTCCGCGGACCCA	2801540
QY	849	1uGluLeuAlaValAlaSprthArgLySAlaGluGluProValSeraGluCySerthArg	869
Db	2801541	CCCAAGCCGCGT---GCCGCGTGGCCACCGGCGCGCGCGTGCCTGCATCCAAAG	2801597
QY	869	1uAlaGluGluGluProAlaLySglYLySAspAlaGluAlaGluAlaThraIaGluG	889
Db	2801599	CMGCGGTGCGCTCCCAACCGGACCGGCGTTCGCGCGTCCGCTCAACCGCGGCG	2801657
QY	889	1yAlaLeuLySAlaGluLySglYglYglYSerGlyArglaAThr-----T	905
Db	2801658	GCGCC-----GAGCGCGGCTTGGCGTCCGCGCGCCACTTGTGGAACGMAA	2801702
QY	905	hAlaLySerSerGlyAlaProGluAAspSerAspSerSeraIaThrCySerAlaAapG	925
Db	2801703	CGCGCTGTGCGCGGAGGCGCCACCGCGTCCGCGCGTCACTTCGCTCGCGCA	2801762
QY	925	1uValAspGluAlaGluGlyGlySAlaSlyAsnArgLeuLeuSer-ProArgProSerLeu	944
Db	2801763	GCGTC---TCCGTCAAAATCCCGGAGATGATGMAAGCCGTACCGCCAAACCAACNTA	2801819
QY	945	1euthrProthrGlyAspProArglaAAsnAlaSerProGluLySProLeuAspLeuLyS	964
Db	2801820	CCAGCGTCCCGCGGACCGCGGTCAACCGCGTCCCGCCCAACGCT-----CGA	2801870
QY	965	GluLeuLySgluArglaAlaAlaAlaIleProPollegluValThrLyValHisGluPro	984
Db	2801871	TTCGCGACTCGCGCGGCGGTCCCAACCGCG-----GCCCG	2801909
QY	985	ProArgGluAspAlaAlaProThrLySProAlaProProAlaProProProGluAsn	1004
Db	2801910	CGGTTTCGCGCGGCGCGCCCAACCGCGGTAGCAACGTGCGCGCGCGCGCGTGC	2801965
QY	1005	-----LeuGluProGluSeraSpAla	1011
Db	2801970	CGGCGCGGTTCGTTAATGTCAAAAGCGGTACACGCGGTACCGCGACCCACCAACG	2802029
QY	1012	ProGluGluProGlySeraSerProArgLySArg-----SerProAlaPro	1028
Db	2802030	CGGCTTAGGCTTCGCGCCCGCGGTCAACACCGTCCGCGTGTAGTTCCGCGGCGCA	2802089
QY	1029	ProAla-----AspLySgluAlaPhe	1035
Db	2802090	CGGCGCGCGCGTCCGCTCCGCGCCACCGCTGCGCGCGTGCAGAACCGTCAAGCCCTT	2802149
QY	1036	AlaAlaGluAlaGluLySLeuProGlyAspPro-----ProCySTrpThSer	1051
Db	2802150	AGGTTCGAGTTCGGGACCAACCGCGGACCGCGCGCGCTTCCTCCGCGCG--	2802206
QY	1052	GlyLeuProPheProValProProArgGluValIleLySAlaSerProHisAlaProsp	1071
Db	2802207	-----CCGCTTCGCGCGCGCGCTTATCCGCGGTGCGCGCGCGCGCGCGCG	2802254
QY	1072	ProSerAlaPheSerYrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThr	1091
Db	2802255	-----TGTCCACCAACCGCGCGCGCGCGCGCGCGCGGCGT-----	2802287
QY	1092	AlaArgProValLeuProArgProProThrIleSerAsnProProPro-LeuIleSerSe	1111

Db	2802288	-----CGCGCCGCGCCAGCCCGACCTTGG-----CGCGCGGCTCGCGCGTTC	2802329
Qy	1111	TAATATyAProSeRValLeuGlnAArgGlnIleGlyAlaIleSerGlnGlyMetSerVa	1131
Db	2802330	CGCGCGTGGCCAGAAATAGCTCGCGCGCGCC-----	2802360
Qy	1131	IGlnLeuHiVaIProGlySerGlnHiAlaIlySaIaProVaIGIProVaIThmEtGI	1151
Db	2802361	-----CACCAAGCCCCACCGCGCCACCA-GCCCCACCTTGGCGGCAGCAAGGGAG	2802412
Qy	1151	YLeuPro---LeuProMetAaPProLybLybLeuAlaProPheSerGIyValysGlnGI	1170
Db	2802413	CCCTTCAGAGGACCGGTGGCGCGCGTGGCCACCTTGGCCCGGCA-----	2802458
Qy	1170	UGlnLeuSeRProAArgGlyGlnAlaGlyIProProGlnSerLeuGIyValProThAlaGI	1190
Db	2802459	-CGCGCGGTACAGATGTGCGCGGGGTACCGGCC-----TTGGCTTCAGAGCCCC	2802505
Qy	1190	nGlnAlaSeRValLeuAArgGlyThraIaLeuGIySeRValProGIyGIySerIleThrLy	1210
Db	2802506	ACCGCGCC-CCCATACCGCCGAGAGCTTGGAGGGGTGGCGGTGCGCCCATACAC	2802564
Qy	1210	SGIYIleProSeRThraRValProSeRaPSeRaIleThrTyAArgISeRleTh	1230
Db	2802565	CGGCTCGCGCTCCCGCAATAGCCCGGCGTGGCGCGGTACCGCCCTGGCGCGCGTGC	2802624
Qy	1230	rHiSGIYThrProAlaAaRValLeuTyLySGIYThrIleThraRGIleIleGIyIuaS	1250
Db	2802625	CGCGCGCGCGCGCGGCG-----CTTGGCACCGTTCG-----	2802655
Qy	1250	pSeRProSeRArgLeuAaRValGIyAArgIuaRSeRleuProLybGIyHiSaIleTy	1270
Db	2802656	-GCCAGCGCCCGCTCGCGCGCGGTGAGCGCCGATTTGCTGGCGGTCAAGCGCGT--	2802712
Qy	1270	rGIuGIYLybysGIyHiSaIleSeRtyrGIuGIyGIyMeSeRValThrGlnCySe	1290
Db	2802713	---GCCCGCAAGGTGCGTTGGCGGTGCACCCCGAAGACCGCGAGCGCGTGGC	2802768
Qy	1290	rLyGIuAaRPIyAArgSeRSeRGIyProPnHiSGIYThraIaIaProLybATGTh	1310
Db	2802769	C-----GGGGTACCGCGCGTTACCGCGCGCTC	2802795
Qy	1310	rTyAaRMeRMeRGIyAArgValGIyAArgAlaIleSeSeRValaSerIleGIyLe	1330
Db	2802796	CGCGCGCGCCCGACGGTGTATACCAATGCCCGCGT--GCCCGCGCGCCCGCAAGC	2802852
Qy	1330	uMeRGIyAArgAlaIleProProGIuAaRHiSeRProHiHiSaIySeuysGIuGlnHi	1350
Db	2802853	CGCGCGCGCCCGCA-GTCCCGCGTGCACACCGCACCGGTGCGTGTCCGA-	2802905
Qy	1350	aIleAArgLySeRleThrGlnGIyIleProAArgSeRtyrValGIuAaRGIuAaRty	1370
Db	2802910	-----CCCTCCCGAAGTGGCGCGCGCCGACCCCA	2802941
Qy	1370	rLeuAArgGIuAaRlybLybLeuLybAArgGIuGIyThrProProProProProSe	1390
Db	2802942	CGGTCCCGCATGTGCACCGACCGCGCGGACCG--CAATCGCGCGCACACCGCGCGC	2803000
Qy	1390	rAArgAaRleuThrGIuAaTyLybThrGlnAlaLeuGIyProleuLybLeuysProAl	1410
Db	2803001	CCCAACGGGTGCGCGCATATCTGTGATA---CCGTGGCGCGCTG-----	2803043
Qy	1410	aHiSGIyLeuValaIaThrValLybGIuAlaGIyAArgSeRleHiSGIuIleProA	1430
Db	2803044	-----GCCCGCGC	2803055
Qy	1430	gGIuGIuLeuAaRHiSThrProGIuLeuProleuAlaProAArgProleuLySGIyI-	1449
Db	2803052	CCCAACGGTATAGGCCACACCGCGGTGGCGCGCGCGCG--CGTTGGC-GCGGGAC	2803107
Qy	1450	-SerIleThrGlnGIyThrProleuLytyAaRThrGIyAlaSeRThrGIySeRty	1469

Db	2803108	CGCCGTCACCGCACCGACCC-----ACGGCGCCACCC-----	2803141
Qy	1469	SLYENIAPVALARGSERLEUILEGLYSERPROGLYARGTHR-PHEPROVALINHP	1489
Db	2803142	-----GGACACCGGACACCGGACCGCGCGCTGCC	2803176
Qy	1469	ROLEMAPVALMETALAAPRLAARGLALEUGIARGLACRYTGLUGLUSERLEUL	1509
Db	2803177	CC-----	2803178
Qy	1509	YSEARARGPROGLYTHRILASERSESERGLYGLYSERLEALARGLYALAPROVALI	1529
Db	2803179	---GGCTCGCCCTGACCCCGCTTGCCTCCCGGTGGCCGAGGACGAGGCATCT-----	2803229
Qy	1529	LEVALPROGLIULEUGLYLAFPROARGINSETPROLEUTHYTRGLIAPRHSIGLYAP	1549
Db	2803230	-----GAACAGGCGCGCCGAGTAATGGCCCGTTGGCCCGCGCGCATGCGGCAC	2803281
Qy	1549	ROPHEALGLYHIULEUPROARGIYSETPROVALTHMETARG-----GLUPROT	1566
Db	2803282	CA-----TTGCCCGCGCGCCACATGCGCCAGCCGCGGCATATGCGCGGT	2803329
Qy	1566	HPROARGLEUGLUGLISERLEUSESERSELYALASERGLMAPARGLYSEUT	1586
Db	2803330	CGCTTCGTTGCC-----C	2803344
Qy	1586	HRSETRHPROARGIULLEALALYSEPROHISERTHRVALPROGLIHIAPROH	1606
Db	2803345	CGCGCGCGCGCGGACGCTGGCTTGCGACCGTTGCCACCGTTACACCGTTGCGCGCG	2803404
Qy	1606	ISPROLSESPROTYGLIHIULEUENARGLYVALSERGLVALAPLEUTYRARGS	1626
Db	2803405	TGCCG---GGCCCGTGG-----TCGCTGCGTGCACACCGCTG---GCCG	2803443
Qy	1626	ERHISILEPROLEUALAPHERPROTHRSERILEPRO-ARGGLYLEPROLEUAPRLA	1645
Db	2803444	CGCGGTCCTGTCGCTGTCGCGCGGCAAG---CGGCSAGAGCTTCGCGCGCGGCGT	2803500
Qy	1646	ALALAIALATYTRYTYLEUPROARGHISLEUALAPROANPPOTHTRYPROHISLEUTYR	1665
Db	2803501	CACCGCGTCACCGCGCTCCCGGCATAT-CTGCATACACCGCGTCACCGCGCTTCCG	2803559
Qy	1666	PRO-----PROTYRLEULIARGGLYTRYPROAPRTHRALALALEUGIAPN	1681
Db	2803560	CGCGCTCGCGGAGCGCGC-GTACCGCGCAACCGCGGCACACGAGGCT-----	2803609
Qy	1682	ARGGLNTHRIELLEAPRPTYRILETHISERGLNIMETHIENIAPNTHRALATHR	1701
Db	2803609	-----	2803609
Qy	1702	ALAMETALAGINARGLAAPRMELEUARGLYLEUSERPROARGIUSERSETRLEUALA	1721
Db	2803610	-----GCCGACCTACCGCGCGGAGTGGCGCGCGCGCGCGCGCGCATC--	2803651
Qy	1722	LEUAPNTRYRALAAGLYPROARGLYLLEIAPLEUSERGINVALPROHISLEUPRO	1741
Db	2803652	-----ACGCGCGGCGCC-----	2803663
Qy	1742	VALLEUVAL-PROPROTHRPROGLYTHRPROALATHRALAMETAPRGLLEUALTYLE	1761
Db	2803664	---GCCGTACCGCGGCAACCGCGCGACACCGCGCTCGCG-----	2803700
Qy	1761	UPROTHRALAPROGINPROPHESERSEARGHISERSESERPROLEUSERPROGLYGLI	1781
Db	2803701	---CTTTCCCGCGCGGCTCCAGAGGCTCAGCGCTGTGGCGCGGAGAGGCGCC-----	2803754
Qy	1781	YPROTHHISLEUTHIRLYSPROTHRTHRTHRSERSESERGLIARGIARGAPARGAS	1801
Db	2803755	-----TTGCGCGCGCGGCC-ACCACATAGCTCGCGC-----	2803783
Qy	1801	PARGGLIARGAPARGAPARGIARGIULYSERLILEUTHSERTHRTHRVA	1821
Db	2803784	-----ACCAACCGACCGCGCGCGCGCGCTTGCACCGGATCCACCGCTCGC	2803830

QY	1821	GLuHISaIaProlleTpaRgProGlyThrGluGlnSerSerGlySerSerGlySerG	1841
Db	2803831	CCGACACCGACCC-----ACGGCGCCACCGGCTCCGC	2803863
QY	1841	rgLyGlyGlyGlySerSerSerArgProAlaSerHis-----	1854
Db	2803864	C-----GGCACCGCGCGCGCGC-CGGCCCTGCCGATGCGCTCCACCGCCGTTGCCG	2803916
QY	1855	-----SerHisAlaHisGlnHisSerProIleSerProArg----	1866
Db	2803917	CCACTACCTTAACGCGAATCCCGGCACACCGCTCACCGCGGACCGCCCGCGCGCCA	2803978
QY	1867	-ThrGlnSerAlaLeuGlnGlnArg---ProSerValLeuHisAlaThrGlyMetLeuG	1885
Db	2803977	CCGCGACCGCGCGCTTACCGCCCGTGGCCCTTCCGCTTGGCCCGGTGGCCGAGAGG	2804038
QY	1885	YIleIleThrAlaValGluProSerLysProThrValLeuArgSerHisSerHisSerS	1905
Db	2804037	GTGCCG-TCGGCG-----CCGTCCTCACTTAACGTC-----GC	2804066
QY	1905	rProValArgProAlaAlaThrPheProProAlaThrHisGlySerProLeuGlyGlyThrLe	1925
Db	2804069	GCCATTGCGACCGCGCGCGCGCTACACACCGTCGCGCGCGCGCGCGCTGTCACCGACCA	2804128
QY	1925	uAspGlyValTyProThrLeuMetGluProValLeuLeuPro-----LysGluAl	1942
Db	2804129	CCCGCGCTTACACCGCGACCGCGCCCTTCCGCGCTTGGCGGACACACGCGGGGTGGC	2804188
QY	1942	aProArgValAlaArgProGlu-----ArgProArgAlaSerPth	1955
Db	2804189	GCCGTACCGCGCTTCCGCGCGCTTGCACACCTGTGCACCGGTACGCGCGGTGCGCGGT	2804248
QY	1955	rgLHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerP	1975
Db	2804249	CACACCCCTGGGTGGCGCGCGCGCGCGCGCGC-----CACCTTCCAC	2804293
QY	1975	oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl	1995
Db	2804294	GCGCGTGGGTACACACCATGCGCGCGCGCGCCACCG-----TCACATCTTGAAGC	2804347
QY	1995	aArgThrProAlaLysAlaLeuAlaProHisAlaHisAlaSerProAspProAlaProP	2015
Db	2804348	CATGTGCGCGCGCGACCGCGCTGCCCGCTTACCGGGGCGC-----CCACACGCGCGCG	2804401
QY	2015	oAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnG	2035
Db	2804402	GGCACACACGTCGCCCGCA-----ACACCTCG-----	2804429
QY	2035	uLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluP	2055
Db	2804430	-CTGCCCGCGCA-----CCCCAGGTGGTGTCAC	2804458
QY	2055	oValSerProValSerSerProSerLeuThrHisAlaPheGlyLeuProLysHisLeuG	2075
Db	2804459	GCCAGCGCCACCGCGCCACCGTACCGCGCGGACCGCGCGCTCCGCCATCA-----CC	2804512
QY	2075	uGluLeuAspLysSerHisLeuGlnGlyGluLeuArgProLysGlnProGlyProValL	2095
Db	2804513	ACCGTTGCGCGCGTCAGATACAGACACAGCATTTGAAACCGTAGCTCCGTTACCA-----	2804568
QY	2095	sLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe	2115
Db	2804568	-----CGGCGCGCGCGCGCCACCGCTTCCGCGGACACCGCGCGCGCGCCATC	2804617
QY	2115	r-----SerSerProLeuLeuGlnThrAl	2123
Db	2804618	GCGGCGGTGGGCGCCACCGCGCGCGCGCGCCACCGCGCGCGCGCTTACCTCATCTCTC	2804677
QY	2123	aProGlyValLysGlyHisGlnArgValValThrLeu-----AlaGlnHisIleSerG	2141
Db	2804678	ACCGGGGGTACCGGATGACCCAGAGAAAGTCCGCTATATCGCATAGCCGACCGACCGCG	2804737

QY 2141 uValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro-- 2160
Db 2804738 GGCTCCGTACACCCATGACCGCGGCCCGCCACCGCATCCGACCGCGCGTGGCCACC 2804797
QY 2161 ----AlaProLeuTyrSerSerProGlnAlaSer-----CyseProValLe 2174
Db 2804798 GGCACCGCGCGTTACCGCCATGACCGCGCGCGCGCGTGGCGCCATTGCGACCTTT 2804857
QY 2174 uAspLeuArgArgProProSerSerAspLeuTyrLeuPro-ProProAspHisGlyAlaProA 2194
Db 2804858 GCCCGCGTTGCCCGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2804916
QY 2194 lAArgGlySerProHisSerSerGlnGlyGlyLysArgSerProGlnProAsnLysThrSery 2214
Db 2804917 CGCGACACCG 2804969
QY 2214 alLeuGlyGlyGlyGlyLysArgGlyLysLeuProValSerProProGlnGlyMetThrGlnP 2234
Db 2804970 -----CCCGGCGTCCGACCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805000
QY 2234 roGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlnThrGlnP 2254
Db 2805001 CGTCACCG 2805045
QY 2254 roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaAspPheSerL 2274
Db 2805046 CGTCACCG 2805105
QY 2274 yLeuThrGlnSerAsnSerAla-MetValLysSerLysGlnGlnLysLeuLysLys 2293
Db 2805106 CGACATACCCCAATGATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805165
QY 2294 LeuAsnThrHis-----AsnArgAsnGlnProGlnTyrAsn 2305
Db 2805166 TTCGCCGCGACCGCATTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805225
QY 2306 lLeSerGlnProGlyThrGlnLysPhe-AsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db 2805226 -----CCGCGCCATTATTCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805273
QY 2325 tThrTyrArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGlnLysAlaI 2345
Db 2805274 CACCGGACACCG 2805328
QY 2345 eIleArgLysAlaLeuMetGlyLysTyrArgGlnTyrGlnGlnLysSerProProLeuSerAl 2365
Db 2805329 -----GAGAAGCCACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805351
QY 2365 aAsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaAlaMetProIleTh 2384
Db 2805352 CGTTGCCCG 2805411
QY 2384 rAlaAlaAspGlyArgSerSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaL 2404
Db 2805412 CGCCGCGC-----ACCGTCGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805453
QY 2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerG 2424
Db 2805454 CTGACCGCGCGCGTACGACCGTTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805513
QY 2424 y-----AspArgProProSerSerValSerSerValHisSerGlnGlyAspCysAsnArgAr 2442
Db 2805514 CCCCATGACCAACCAACCG 2805570
QY 2442 gThrProLeuThrAsnArgVal-TripGlnAspArgProSerSerAlaGlySerThrProP 2462
Db 2805571 CACCGGACACCG 2805630
QY 2462 hPro-----TyrAsnProLeuI 2468
Db 2805631 CGCGCACACGACCG 2805690
QY 2468 lMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlyS 2488

Db 2805691 TCCCGCGGTACTCCGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2805750
QY 2488 ergGlyProLeuAlaGlyPro 2494
Db 2805751 ATACCG 2805770

RESULT 14

US-08-804-227C-1/C
Sequence 1, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kunstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1
Alignment Scores:
pred. No.: 1,25e-20 Length: 43280
Score: 661.50 Matches: 656
Percent Similarity: 28.75% Conservative: 275
Best Local Similarity: 20.61% Mismatches: 1111
Query Match: 5.01% Indels: 1310
Gaps: 138
DB: 2

US-09-522-753-5 (1-2517) x US-08-804-227C-1 (1-43280)

[illegible]

QY	392	LeuAlaValIleProPheMetLeuTyrAspAlaAspGlnGlnArgIleLeuPheIleAsn	411
Db	12135	-----	1213
QY	412	MetAsnGlyLeuMetAlaAspProMetLeuValTyrTrpAspArgGlnValMetAsnMet	431
Db	12134	-----	1211
QY	432	TrpSerGlnGlnGlySerGlnThrPheArgGlyLeuPheMetGlnIleProTyrAsnPhe	451
Db	12110	CGTCAGCCGCGACCCGCGAAGACACCGGACCCGTCGCGTAAGACACACTCC-----	1206
QY	452	GlyLeuIleAlaSerPheLeuGlnArgTrpThrValIleGluCysValLeuTyrTyr	471
Db	12059	-----	1201
QY	472	LeuThrTrpAlaAsnGluAsnTyrTyrLeuSerLeuValArgArgSerTyrArgArg	489
Db	12017	-----	1196
QY	490	-----	501
Db	11963	CACACCGAACCGGCGCGCACCGGTCAGAGCCCGGTCCGGCAGAGACACTCCGCGACAG	1190
QY	502	-----	510
Db	11903	CGCGTCGGCGGTGCGCGCATACGGCGCGCACCGCGACGCGCT-CACGTAATCTTAC	1184
QY	511	ProMetProArgSerSerGlnGlnGlnLeuLeuAspGlySerGlyLeuGlySerGlnAlaGlu	530
Db	11844	CCAGAGGCCCGCGAACCGTTCTGTGAAGACTCGGCGGTTTGAAGCGTTTCCCGCGCGCGCC	1178
QY	531	Leu-----	538
Db	11784	AGCGCCGCGCCCACTCGGCGCTCCGAGAGCGGTGGGCTCCGAGACAGCGTGGCT	1172
QY	539	-----	552
Db	11724	CGCGCTGCGCGTCCAGACACCGGCGAGTCCGCGCTTCTGCGCGCGAATGACGCGCG	1166
QY	553	ThrSerGlyLeuAspAsnAspGlyLeuGlnAlaValAlaSerTyrSerGlyTrpThrAla	572
Db	11664	CTTCCGGGCGCGAATGATGAGCACTCCGCGCTCCGACTGTGCGCGCGACACCC	1160
QY	573	AsnSer-----	583
Db	11604	TCAATCTACCCGCGACACCGCGCGCGGCGAAGACACAGCGCGCTCAACGCTACGT	1154
QY	584	-----	595
Db	11544	CTTCCAGCGCGCGAGCGGTGGCGCT-CCGCGTGCAGCGCGAACTCAGAAAGGCC	1148
QY	596	-----	603
Db	11485	GTAACCCGCGACAACTCAACCCGCGACCGGCAATGGTCCGTCAAGCAGTGTCTCGAG	1142
QY	604	LeuAlaSerMetGluLeuAsn-----	614
Db	11425	CCGCGCAGCGCGCGCGGAAACCCGCTCAACCGAATGCCAGGGGTAAACCGCCACGG	1136
QY	615	-----	618
Db	11365	AGGAAGGGGTGTCTCTCCACAGTCAATACGAAACCGGCGCGACCGCGTGGCGACTCGG	1130
QY	618	-----	618
Db	11305	GCGGCGAGTGTACAGCGCCAGTATGTGTGTGTAAGGGATAGGTGGAGGATGTGTG	1124
QY	618	-----	618
Db	11245	AGGTAACTGGGGGACAGGCGGGGCTCCACAGTAAAGGATGGCGCGGTGGCCACAGGCTTG	1118
QY	619	GluMetGluThrAlaLeuValGlyLeuLeuGlnIleGlyArgAsnTrpSerAlaIleAla	638

```

Db      11185 GCGGTGAGAGTGCAGGCGGCGGCTGCTCTGCGGCGAGGCTTCGGTCCGAGTCA 11126
QY      639 ArgMetValGlySerThrValSerGlnCysIysAsnPhenylrhisAntyrlslys 658
Db      11125 TGGACACTCCGGTCCGGGCGCTCGGCGCTGCTC-AATGGCATGAGCAAGTACCGGGTG 11067
QY      659 ArgGlnAsnLeuAspGlu-----IleLeuGlnGln-HisIysLeuIysMetG1 674
Db      11066 GGCATCTGACCTCGATGACCGACGCGTGCGCTCTCCAAACAAACACCGACCAACCGCGGA 11007
QY      674 uIysGluAAsnAlaArgArgIlysbysbysAlaProAlaAlaAlaSerGlnGluAl 694
Db      11006 GAATCTAACCCGGTTCGAGATTACGAACAATACCGCATCGAAAAACCGCTCGCC 10947
QY      694 aAlaPheProProValValGluAspGluMetGluAlaSerGlyValSerGlyYasnG1 714
Db      10946 CGGCTGCTCTCCGGGCGACGGTG---GAACACACCGGACCCCGGAGCTCGGGGGGGGAT 10890
QY      714 uGluGluMetValGluGluAlaGluAlaLeuHisAlaSer----- 727
Db      10889 ACCGGACACACCGGCTCCCACTCGCCCTTCAGGTCTCTCAAGTCGAGGAGTGAAGGC 10830
QY      728 ----GlyAsnGluValProArgGlyGlyIucysSerGly----- 738
Db      10829 GTAGTCCACCGGTATCAACCGAGCC---TGACACACCTCCGCTCGACAGCGGCACATA 10773
QY      739 -ProAlaThrValAsnAsnSerSerSerAspThr----- 748
Db      10772 ACCGGCACCGCGCGCGCATCCCGGAAACACCGGTGACCGCGACCGGTACCGCGGC 10713
QY      749 -----GluSer-IleProSerProHisThrGluAlaIlyb- 760
Db      10712 TACTGTACTCGGGCCACTTCGCGACCGCGGCTCGACTCGCGGAGCGAGCGGAAC 10653
QY      761 -----AspThrGlyGluAsn----- 765
Db      10652 AGCCGCGCATCGACCAACGACCCGCGCATACCGGCATCAACCCGCGCGAGACGAC 10593
QY      766 ----GlyProIysProAlaThrLeu----- 773
Db      10592 CACAGCGCGCGCATCTCCAGGCAACAGCCCGCGCACCGTGGCAGCAAGCATCTCCC 10533
QY      774 --GlyAlaAspGlyProPro-----ProGlyProP 783
Db      10532 CTGGGAATACCGCACCGACGCGACGTCACACCATCGCTGCGACGTACCGGGCGCAG 10473
QY      783 roThrProProArgArgThrSerArgAla----- 792
Db      10472 CGACACCATACCGCGCGCGTCACGGGCTGGCACACATCCACCGCTCAACCCGCGCCC 10413
QY      793 -----ProIleGluProThrPro----- 798
Db      10412 GCCCGACACCACTCCAGCAAAATCCCACTCCACATGAACATCAGGCGCGGCGACATC 10353
QY      798 ----- 798
Db      10352 CCGCATCGACGCGCAACACCTCCGAAGACCCAGACGCGCGCACCATCTCCCAACCA 10293
QY      799 --AlaSerGluAlaThrGlyAlaProThrPro-----ProProAlaPro- 812
Db      10292 CTGGCTCCCTTGACAGGAAACAAACACGACCGCCCAACCGGACCCGCTCTCCGA 10233
QY      813 -----ProSerProSerAlaProProProVal----- 821
Db      10232 GTTCACACACCGCACCGGTGACACACCCGCTCGGATCAACCCCGCGCAACCGCGCAA 10173
QY      822 -----ValProIysGluGluIysGluGluGluThrAla 833
Db      10172 AGCGCCGCGAAGTCAAGCTCAACCCGAGCGACACAAACCGCGCGGTACCGCAACCGCGCA 10113
QY      833 laAlaProProValGluGluGluGluGluGluIlybProProAlaAlaGluGluLeuAlay 853

```

```

Db      10112 CCGGCGCGCGCA-----TCGTACCGCGCACATCCCGCAGCGCCCGCACACCG-GCC- 10064
QY      853 aAlaPThrGlyIlybAlaGluGluProValIlySerGluCysThr-----GluGluAlaG 871
Db      10063 --TCACACCATCTCCCGCAACCGCGCGCGCTCCCGCACACCGCGCAACCGCGCG 10006
QY      871 luGluGlyProAlaIlybGlybysbysAlaGluAlaGluAlaThrAlaGluGluAla 891
Db      10005 AAACCAACACCGCGCACACCAACCGAGCGTGTGCTCTTCGGAGATCGCAGAGAGCTT 9946
QY      891 eu-----LysAlaGluIlysbysbysGluGlyIlySerGlyYArgAlaThrThrAlaIlyssers 909
Db      9945 TCGCACACCTCCGCTCGGTGGGTGGGTGGGGCTTCAGAGGCTTCGGTGAAGCTGTGGAGCT 9886
QY      909 eGlyAlaProGluAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluA 929
Db      9885 CGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGAGCTTCGAGGCTTCGAGGG 9826
QY      923 laGluGlyGlyAspIlybAsnArg---LeuLeuSerProArgProSerLeuLeuThrProT 948
Db      9825 CTTCGGGGGCTTCGAGATCAAGATGAGCGGTGTGACCA-----CTGACACCGCA 9778
QY      948 hrGlyAspProArgAlaAsnAlaSerProGluIlybProLeuAspLeuIysGlnLeuIysG 968
Db      9777 A-GCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTC 9719
QY      968 laArg-----AlaAlaAlaIle----- 973
Db      9718 AGCAACCGCACACACGACGACCGACGATCAACATCCGCTCGGTCAATCGACATCAAC 9659
QY      974 -----ProIleGlnValThrIlybVal-----HisGluProPro- 985
Db      9658 GTCCGCGCAACCCACCATACC-GCATGCGCAACACCATCTTGATCAACACGACCAACCC 9600
QY      985 ----- 985
Db      9599 CCGACCGCGCTGAGCATGACCAATATTCGACTTCAACGACCCCAACACACCGGACATTC 9540
QY      986 --ArgGluAspAlaAlaProThrIlybProAlaProProAlaPro----- 999
Db      9539 AGCATCAGATCCCGCCCATAGCTCCCGACGAAACCTTCGCTCGATCGATCGATCAACCGAG 9480
QY      1000 -----ProProProGluAsnLeuGlnProGluSerAspAlaProG 1013
Db      9479 CCGGCTCCCGTCCATGCGCTTCACCAACATCAATCCGCGACCGTCAACCGG---- 9424
QY      1013 lngIlnProGlySerSerProArgIlyIlySerArgSerProAlaProProAlaAspIlysg 1033
Db      9423 -----CATTCGCCAAG---CTGAGAGATCAACCGCTCTCTGCGAGGACCACT 9378
QY      1033 luAlaPheAlaAlaGluAlaGlnIlybLeuProGlyAspProProCysTrpThrSerGlyL 1053
Db      9377 CGGCGCGGTCAAAACATTCGACGCGACCGCTGATTCACCGCGGACCCACACA----- 9325
QY      1053 euProPheProValProProAlaGluValIleIlybAlaSerProHisAlaProAspProS 1073
Db      9324 --CCACCGCGCAACCGGAT-----GCCCAACCGCGGCGCATCCGA 9285
QY      1073 eAlaPheSerThrAlaProProGluIlybProLeuProLeuGlyLeuHisAspThrAlaAla 1093
Db      9284 CAACCGCTCCACACGACACACCGCGACCTTCGCGCGCAAGGCTCCAT----- 9235
QY      1093 rgProValLeuProArgProProThrThrIleSerAsnProProProLeuIleSerSerAlaL 1113
Db      9234 --CGCGCCCATCGGAAAGCTTCACCGCGCATCGCGCGCAACCGCGCGCGCGCA 9177
QY      1113 yHisProSerValLeuGluAlaGlnIleGlyAlaIleSerGlnIlyMetSerValGlnL 1133
Db      9176 AAATCTCAAGA-----ACAATCGGGGTGCGCGCATCAACGCTCAC 9138
QY      1133 euHisValProThrSerGlnHisAlaIlybAlaProValGlyProValThrMetGlyLeuP 1153
Db      9137 ACCACCGCGCAACGCGCATCTCACTCGCGCACCGCGCAACGCGCTCGG----- 9091

```

Qy 1153 roleupromewaspProlyseVleuAlaproheserGlyVallyeGlnGlnLeus 1173
Db 9090 -----CCGGTTGTGTACAGGGCGACACGACGACGACGACGCGGTCTC 9048
Qy 1173 erProAlaglyGlnAlaGlyPro-ProGlnserLeuGlyValProthralaGlnAla 1192
Db 9047 GATGGAAGAGGAGGACCTTCACAGCCGCTACAGATGCGCGGAGAGAGAGCGCT 8988
Qy 1193 serValleuArglyThralaLeuGlySerVal-----ProGly 1205
Db 8987 CGCGACCTTGGCGGTGACAGATAGCTTCTACTCTCTCGGGGCTCTGTCAGCCGGA 8928
Qy 1206 GlySerIleThrylsglyIleProserThxArgValProserAserAlaIleThrTyr 1225
Db 8927 GGGTACTCTCGATGGGTACGCCGAC-----GTATAC 8895
Qy 1226 ArgGlySerIleThrIhsglyThrProAlaAserValLeuTyrlyGlyThrIleThrArg 1245
Db 8894 CCGCGTCTCG-----GCTGCCGTGACAGCA 8871
Qy 1246 IleIleGlyGlnAserProserArgLeuAserArglyValArgIleAserLeuProlys 1265
Db 8870 GTC-----CGGGTCATACCGCGCGCTCGATCGCTCCACGA 8832
Qy 1266 Gly-----HisVal-IleTyrGlyGlylyValysGlyHisValIle 1278
Db 8831 GGTCTCAGACACGCGCGCTGCTCGGATCATTCACGCGCTCAGCGCGGAAATCCC 8772
Qy 1278 userTyrGlyGlyGlyMetSerValThrGlnCySerIleGlyAserIleArgser-Sers 1298
Db 8771 GAAGAATCCCGCTGAACTCCGCGCGCTGAGAAACCCCGCTCCGACATATGT 8712
Qy 1298 erGlyProProhIhsglyThralaAlaProlyAargThrTyrAserMetMetGlyArgv 1318
Db 8711 CTTCGCCGACACCGCGCTCGGGCTGTAGACGGAATCCAGGTCCAGCGCGGTGCT 8652
Qy 1318 alGlyArgAlaIle-----SerSerAlaSerIle----- 1327
Db 8651 GGGCGCGCGCTATGACGTACGCTCTCGGTACCAAGCTTCACAGCGCGCTCAGGGT 8592
Qy 1328 -----GlnGlyLeuMetGlyArgAlaIleProProGlnArgHisSerP 1342
Db 8591 GCCCGCCACCGGGGAAAGGCGACGCGCATGCGGACGATGACCGCGGTGCTGCGG 8532
Qy 1342 rohIhIhVleuIyVgIuGlnIhIhIleArgGlySerIleThrGlnGlyIleProArgS 1362
Db 8531 CACCG----- 8527
Qy 1362 erTyrValGlnAlaGlnIuAserTyrLeuArgAlaIleValyLeuLeuValArgGlyG 1382
Db 8526 -----CGACGGGACGTGGCGGCGCTCATCGCGCGG-G 8497
Qy 1382 IyThrProProProProProProSerArgAserLeuThrGlnAlaIleTyrlyThrGlnAla 1402
Db 8496 GCCCGCGCGCTCGCGCGCTGACGCGGTGACAGCACTGCGGCGCGGTGCGGCG 8437
Qy 1402 euGlyProleuIyVleuIyProAlaIhIhIleGlyLeuValAlaThrValIyGlnAlaG 1422
Db 8436 TGGGGTGTCTGAAGATG-----ACGCTGGCGGTGACT 8404
Qy 1422 IyArgSerIleIhIhIleProArg-----GlnGlyLeuAargHisIleThrProG 1438
Db 8403 TCAGGCTGTGCGCGCATGAGCGGTGCGGAGTTCAGAGAGCTCAGGAGTCAAAAC 8344
Qy 1438 IuLeuProleuAlaProArgProleuIyGlnGlySerIle-----Thrg 1453
Db 8343 CGCGCACTTGAACGCGCGCTTCGCGCGCATGCGCTCTGTTCTCGATGCGCGGACGCA 8284
Qy 1453 IuGlyThrProleuIyTyrAserThrGlyAlaSerThrThrGlySerIyLeuHisAser 1473
Db 8283 AGGCGCGCTCGGTACG-----ACGAGTTCGACCACTTCGCGGCGGTTCGG 8236

Qy 1473 al-----ArgSerLeuIleGlySerProGlyArgThrPhePro 1486
Db 8235 TCTCGTTACAGATGACCAACCGCGCGCGCTCGCGCGCGGTCTCGCGGCTC----- 8183
Qy 1486 roValhIhIhProleuAserValMetAlaAserAlaArgAlaLeuGlnAlaCyTrGlyG 1506
Db 8182 -----CGGACGTGTGTCTCGCTCGTACAGGGCTCTCGCGCTCGCGGACG 8134
Qy 1506 IuSerLeuIySserArgProGly----- 1513
Db 8133 TGTCTG---AACAGCGCGCGGACGATGCGCTTCTGCGCGCGGCAACGCTCCCACT 8077
Qy 1514 -----Thralas 1516
Db 8076 CGAAGTCAGACAGATGACGACACGCTGTTCTGCGCCATGGGCGCCAGAGCGGTCTGA 8017
Qy 1516 erSerSerGlyIySerIleAlaArgGly-----Alap 1527
Db 8016 CGGCGCGGTGCGGTGCTCATGCGCGCGCGCGCGCGCGCATCTCTCGCGCGCAC 7957
Qy 1527 roValIleValProGlnLeuGlyIyProArgIyIserProleuThrTyrGlnAserHisG 1547
Db 7956 CCGCGCGCATGCT-----CCCCCGCCAAAGGCCGAGCGAGGTGCGG 7906
Qy 1547 IyAlaProPheAla----- 1551
Db 7905 GCAATCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGGTGCGCGCGGT 7846
Qy 1552 -----GlyHisLeuProArgIySerProValThrMetArgGlyP 1565
Db 7845 ACCGACCTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 7786
Qy 1565 roThrProArgIeGlnGlnGlySerLeuSerSerIyValAserGlnAserGlyIyS 1585
Db 7785 CCAACCTTTTATGTCGCGATGATGTTGTCAGCAAGCTTCGCGCGCGAC----- 7736
Qy 1585 euThrSerThrProArgIyIleAlaIyS-SerPro----- 1596
Db 7735 --ACCTTGGCGCGCGGAGCGGTGCAAGCTCTCGGTGACACAGCTGTGATACCGCGGT 7678
Qy 1596 ----- 1596
Db 7677 CGTCCAGAAATCCCGCGGTGTGGAAGACGCGCTTCGCGGATACGCGGTGACAGCGCG 7618
Qy 1597 -----HiserThrValProGlnIhIhIhIhIhIhIhIhIhIhIhIhIhIhIhIh 1614
Db 7617 CGAGCGCATGACGCTCGCGCATGACGCGCGGT--GCACGACCTGACGACCATGTC 7561
Qy 1615 Leu----- 1615
Db 7560 CTTCGACGTTCTCGCGGATTCGCGCGCGCGCGCGCGCTCGCGTCCGCGACGCTGTGA 7501
Qy 1616 -----ArgGlyValSerGlyValAserLeuTyrArgSerHisIleProleuAla 1631
Db 7500 GACACGAGGCTTCGCGCT--CGCTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCA 7442
Qy 1632 PheAserPro-----ThrSerIleProArgGlyIleProleuAserAlaIleAlaIa 1648
Db 7441 CCCATGCGCGCGGTATGACAGCGGTGCGAGCGGTGCGAG--GAGGCGGCGCGCGGT 7385
Qy 1649 TyTyrLeuProArgHisIleuAlaProAserProThrTyrProhIhIhIhIhIhIhIh 1668
Db 7384 GCGCGCGCG--CGGGAAC-----CGCGCGCGCGGTAGATCCC----- 7350
Qy 1669 LeuIleArgGlyTyrProAserThralaAlaLeuGlnAserGlnThrIleIleAserA 1688
Db 7349 -----CGAGG----- 7344
Qy 1689 TyTyleThrSerGlnIh 1708
Db 7343 -----CGGACGCGGACGCTGCTGCGCGCGCGCGCGCGCGCGGAGAC 7302
Qy 1709 -----MetLeuArgIyLeuSerProArgIyIuSerSerIleuAlaLeuAserIyAla 1725

```

Db      7301 GCGCGAGCAGAGCTCAAGGCGCGTGGT-TCGAGCTCTGGCGACCTCTCCGGGGTCT 7243
Qy      1726 AlaGly-----ProArgLysLeuLeuSer 1735
Db      7242 GTGGGGGCTGTGAGTCTCCGGCTCCGGTCAAGCTCTGGTCTCCGGCAGATTCATCA 7183
Qy      1736 GlnValProHisLeuPro--ValLeuValProProThrProGlyThrProAlaThrAla 1754
Db      7182 AGCGCGCCCAAGATATCCGGGAGTTCAAGGGCGCGCATCGTCCGACGACCCCAACATGGG 7123
Qy      1755 MetAspArgLeuAlaArgLeuProThrAlaProGlnProPheSerArgHisSerSer 1774
Db      7122 CCCCCCGCTCTGGCGCTCTCCCGGAAACACGCGCACGCGCTCCGGTCAAGCGCC 7063
Qy      1775 SerProLeuSerProGly-----ThrLeuWe 1780
Db      7062 AGACCTTGGCGCGCGGAGACCGCGATGTCCGAGACCCCTGGAGAGCGCTATGACGTCC 7003
Qy      1781 -----GlyProThr 1783
Db      7002 TGACCGCGTCCGTGAGAGGACACGACCGAAGACGAGACGAGTGTCTCTCC 6943
Qy      1784 HisLeuThr----- 1786
Db      6942 CCGCTGACTCATGAGAGGCTTGAGCATTCGCGCATTCGCGACCGCGCGCGCATGG 6883
Qy      1787 -----LysProThrThrSerSerSerGlyArgGln 1797
Db      6882 GCACCTCAAGAGCGCGGACGAGGACACCGCGCTCGAGCAACCGTTGTACGTCGCGCA 6823
Qy      1798 ArgAspArg-----AspArgGluArgAspArg 1806
Db      6822 GAGGCTCGCGCTCGGTCTGGGTGCGGGGTGCAACGAGAGGACGACCGATGAGAGGCG 6763
Qy      1807 AspArgGluArgGlyLysSerLysLeuThrSer--ThrThrThrValGlnHisAlaPro 1825
Db      6762 GAGCGGGAGGATTCGATCAGGCTTTCAGGTGACGGGTAGCGCGACATCGG-- 6706
Qy      1826 IleTrp----- 1827
Db      6705 --TGGGGCTGTGGGTGTCTGTGCGGGGTGTGGGGGTGTGTGTGAGCCAGTAATGGT 6649
Qy      1827 ----- 1827
Db      6648 GGTGTTGAAGGGGTAGTGGGAGGTGTGTGAGTGTGCTGGGAGGACGGGGTCCC 6589
Qy      1828 -----ArgProGlyThrGlnLysSerGlySerSerGly 1839
Db      6588 AGGTGAGGGTGGCGCGGTGGGCCCAAGGCTCGCGGTGGAGTCAACAGCGGTGGCGGC 6529
Qy      1840 SerSerGlyGlyGlyGlySer-----SerSerArgPro 1851
Db      6528 TGTGCTGTGGCGGCGAGGTCTCCGGTGCATGACATCCGGTCCGCGCTCCGCGCG 6469
Qy      1852 AlaSerHisSerHisAla----- 1857
Db      6468 TGTGCTGCATTCGATGAGAGTACCGGGTGGGCACTGACCTGATGAACCGACGGTGGC 6409
Qy      1858 -----HisGln-His-----Se 1861
Db      6408 CCTCTCTCAACAAACCAACGACCGACCGGAGAACTCAACCCGGTTCGCGAGATTACGA 6349
Qy      1861 rProIle----- 1863
Db      6348 ACCAATAACCCCGCATGAACAAACCGGCTCGCGCGTGTCTCCGGGACCGTGAACACA 6289
Qy      1864 ----SerProArgThr-----GlnAspAlaLeuGlnGlnArgProSe 1876
Db      6288 CCGGACACCGCGGACTGCGGGGGCGGATACCGGACGACCGCGCTCAACTCGCCCTTCA 6229
Qy      1876 rValLeuHisAspThrGlyMetLysGlyLysLeuThrAlaValGlnProSer----- 1893

```

```

Db      6228 GGTCTCTCACATGCGGGAGTGAAGGCGTAETCCACCGGATTCACACGAGCTGCACAC 6169
Qy      1894 -----LysProThrValLeuArgSerThr-- 1901
Db      6168 CCTCGGCTGACAGACGAGGACATTAACGGCGACCGCGCGCGGATGCC-CGAAACACAC 6110
Qy      1902 -----SerThrSerSerProValArgTr 1909
Db      6109 GTGACGCGGACCGTTGACCGCGGACCTTCACACCGGCGGACTTCCGCGCGCGCGC 6050
Qy      1909 o-----AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThr 1924
Db      6049 TGACCTGCGCGGACGAGGACGGAACAGCGGCTATGACACGACCGG--CGAGATAC 5993
Qy      1924 rLeuAspGlyValTyPro-----ThrLeuWe 1933
Db      5992 CGGCAATACACCGCGCGGACGAGGACACGACCGCGCGCATCTCCAGGACACACGCG 5933
Qy      1933 rGluProValLeuLeuProLysGlnAlaProArgValAlaArgProGluArgProArg-- 1952
Db      5932 CCGGACACCGTGGCAGACGCGATCTCCCGTGGAAATGACCCACACCGGACGACGCTCC 5873
Qy      1953 -----AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG1 1968
Db      5872 ACAACCATGCGCTGCGATTCAGGCGACGACACCATACCGCCACGTCACGGGCT-- 5815
Qy      5814 ----GACACACATCCACCGCGTCCACCGCGCGCGCGCGCGCGCGACACACATTCACG 5765
Db      1988 Lysr-----GlyHisAlaThrLysAlaArgThrProAlaLysAsn-- 2001
Qy      5764 AATCCACTCCACATGACACTCAGCGCGCGCGGACACTCCGCGATGACGCGCGCAAC 5705
Db      2001 ----- 2001
Qy      5704 ACCTCGAAGACCCACGACGCGCGACCCCATCCCATCGGTCCCGTGCACGACCGAGG 5645
Db      2001 ----- 2001
Qy      5644 AAAACAACACGAGCGCCGACCGGACCGCGTCTCGGGTCCACACGACCGCGGTG 5585
Db      2001 ----- 2001
Qy      5584 ACCACACCGCGTCCGATCACCCCGGCAACGCCCGCAACGCCCGCAAGTCAAGCC 5525
Db      2001 ----- 2001
Qy      5524 TCACCCCGACGACCAACACCGCGGATTCACCAACCGGACCGCGCGCATCTGCACC 5465
Db      2002 -----LeuAlaProHisHisAlaSerProAspProPro-- 2012
Qy      5464 GCCACATCCGCGACCCCAACACCGGCTCACACACTTCGCGCAACCGCGCGGACGC 5405
Db      2013 -----AlaProProAlaSerAlaSerAspPro-----His 2022
Qy      5404 TCCCGCACCCACACCGCGGACCGCGGCAACACACACCGGACACACACTGCGGAGCA 5345
Db      2022 sArg-----GluLysThr-----GlnSerLysProPheSerLysGlnGluGlu 2037
Qy      5344 TCGAGGACCGGACGACACTCCGACACTCAGAGCGGACCGGAGTCTGTGGCGCTCCGCG 5285
Db      2037 uLeuArgSerLeuGlyTyHisGlySerSer----- 2047
Qy      5284 TCGCGGTGTGGGGGCTTCGAGATCAGATGGGGGTGTGACACTGACACCAAGAGCG 5225
Db      2048 -----TySerProGlyValGlnProValSerProValSerSerProSerLys 2064
Qy      5224 GAGACCGCGCGCGACGCGCGCGGTTCACCCCGGCGACTCCCGCGCTCGGTCAACAAC 5165
Db      2064 uThrHisAspLysGlyLeuProLysHisLeu-GlnGluLeuAspLysSerHisLeuGlu 2084
Qy      5164 CACACACACGACCGGACGAGTCCACATCCGCGCTCGGCTC-----ATCCACATGCAA 5112

```

```

Qy 2084 lylguleuAArgProLysGlnProValLysLeuGlyGlyAlaAlaHisLeuP 2104
Db 5111 -----CGTCCGGGGAACCA-----CCCATACCGCATCGCCAAACACATCTTG 5069
Qy 2104 roHlsleuAArgProLysGlnProLysSerSerSerProLeuLysGlnThrAlaP 2124
Db 5068 ATCAACAC-CAGCCCAACCCCGCGCTGAGCATGACCAATATTCGATTCAGCAAGACC 5010
Qy 2124 roglValLysGlyHisGlnArgValThrLeuAlaGlnHisLsSerLysValLleT 2144
Db 5009 CA-----ACCA 5004
Qy 2144 hrglnAerTYrThrArgHisHisArgProGlnGlnLeuSerAlaProLeuProAlaProLeuT 2164
Db 5003 CACCGGACGGCCACCATCAGATCCGCCCATAGCTCCCGACGAAACCGCTCGCTGGAT 4944
Qy 2164 ySerPheProGlyAlaSer-----CyProValLeuAerPheLysArgProProS 2181
Db 4943 CGGATCACCAACCGCGCTCCCGCTCCATGCGCTCCACCAATCCATCCGCGCACCT 4884
Qy 2181 eHsPheLysTYrLeuProProPro-----AspHisGlyLysAlaP 2193
Db 4883 CAACCGCGCATTCCTCCCAACGCTTACGAATCACCGCTCTGCGACGAGCACATCGGCGC 4824
Qy 2193 roAlaArgLysSerProHisSerGlnGlyGly-----LysArgSerProGlnProAlaP 2211
Db 4823 CGTCAACCAATTCAGACCAACCGCTCTGATTCACCGCGGACCCACCAACACCGCCACAC 4764
Qy 2211 yHsThrSerValLeuGlyGlyGlyLysArgGlyLysGlnProValSerProProGln 2229
Db 4763 CGGATGCCCAACCGCGCGCATCCGACCAACCGCTCCACGACGACCAACCGCGGACCTC 4704
Qy 2230 -----GlyMetThrGlnProGlyHisSerArgSerAlaValTYrProLeuLysTYrArg 2248
Db 4703 GGGCCAAACCGGCTCCATCCCGCTCCCGCATCGCGAAC-GCC-----TTGACCGGC 4657
Qy 2248 spglYglGlnThrGlnProSerArgMetGlySerLysSer-----ProGlyLysAlaP 2265
Db 4656 CATCGCGCGGCAACCCCGCTCGCGCAAAATTCAGAAATTCACGAAATTCACCGCGCATCA 4597
Qy 2265 hTsSerGlnProProAlaPhePheSerLysLeuThrGlnSerAsnSerAlaMetValLys 2285
Db 4596 CGGTACACCAACCGCGC-----AACGCGCAT-GTCACACTCG 4562
Qy 2285 eTyLysGlnGlnGlnLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 2299
Db 4561 CCACCCCGCAACGCTGACCGCGCATGCAACGCGCAACGCGCAACGCGCAACGCGGTA 4502
Qy 2300 -----AasGlnProGlnTYrAsnL 2306
Db 4501 TCACCGGTACAGCGCGGACCTCCAGCCCAAGGCTGAGGACACACCGCGGACACCA 4442
Qy 2306 lEsSerGlnProGlyThrGlnLysPheAsnMetProAlaLysThrGlyThrGlyLeuMet 2326
Db 4441 CTCGCCGACGTACCGGTGAGGAGTAGACGTCGCTGCTTCGCTCGGCTCTTGGC-----G 4388
Qy 2326 hTyTyArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGlnAlaLys 2346
Db 4387 AGCCGAGGTCCGATTCCTGCGGATCCACGCGCGGTAGCGCGGCTGCGCTGCG 4328
Qy 2346 lEArgValAlaLeuMetGlyLysTYrArgGlnThrGlnLysProProProLeuSerAlaP 2366
Db 4327 AGCGAGT-GGG-----GTCGATACCGCGC-----GTCGATGCGCTCCAGAGG----- 4282
Qy 2366 snAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProLysThrAlaP 2386
Db 4281 -----TCTCCAGCAACAGCGCGCTGCGGTGATCGATCGACCGCGCTCAGCGCGCA 4230
Qy 2386 lAerGlyArgSerArgPheHisThrLeuThrSerProGlyGlyGlyLysValLys 2406
Db 4229 AATCCGAGAGACTCCGATCGA--ACTCGCGG-----CATCATGACAGAAACC 4182

```

```

Qy 2406 eArgLysArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp 2426
Db 4181 CCCCTCCGCAATAGACTTCCCGGACACCGCGCTCGGGGTGTAACAAC----- 4129
Qy 2426 rGProProSerValSerSerValHisSerGlnGlyAspCyAsnArgArgThrProLeuT 2446
Db 4128 --CCTCCACATCCAGCACGATCACCGGAAACCGGGAAACCGCTGCCACCCCT---- 4075
Qy 2446 hTsAsnArgValTTPGlnAspArgProSerSerLysGlnSerThrProPheProTYrAsn 2466
Db 4074 -----CCACCAACCAACCGCACACGCT--CCGCGAAC 4041
Qy 2466 roLeuLysMetArgLeuGlnAlaGlyValMetAlaSerProPro-----ProProGlyL 2484
Db 4040 CA-----CCCAACCGGGAACCGGACCCCATACCAACGCGATCGGCTCTGCGC 3987
Qy 2484 euProAlaGlySerGlyProLeuAlaGly 2493
Db 3986 TTCGAGCTCTCTGCGCGCGTGCAGGA 3958

```

RESULT 15

US-08-804-227C-13/C

Sequence 13, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.

APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LULY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII (DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804/227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 13987 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 350..13987

US-08-804-227C-13

Alignment Scores:

Pred. No.: 1,01e-20

Score: 652.00

Percent Similarity: 30.21%

Best Local Similarity: 21.56%

Query Match: 4.93%

DB: 2

Length: 13987

Matches: 648

Conservative: 260

Mismatches: 1034

Indels: 1081

Gaps: 133

US-09-522-753-5 (1-2517) x US-08-804-227C-13 (1-13987)

Oy	15	ThrGluProArgTyrProProHis-----SerLeuSer	25
Db	10684	ACCCGACACCGACACACCCCAATCCGCGAAGAACGACCGCTCCCGCGCATAGAGG	10625
Oy	26	TyrProVal-----GlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu	41
Db	10624	CACCCAAATACCCCGGATATCCCAATCTCCCGGAACCCACACCCGACACCCGA	10565
Oy	42	TyrGlnHisSerAspArgSerTyrAlaSerHisLeu-----SerProGlySer	57
Db	10564	CACACACACCAACACCCGCGCGCACCCGACCGAACTCCAAAGCATCTCAACAACCCCGCGAA	10505
Oy	58	IleIleGlnProGlnArgTyr-----ArgProSerLeuLeuSerGluPheGln	73
Db	10504	CTC-----CCCAACACCGGCTTCATCAACACCGAATGAAACCATGACACATCTCAA	10455
Oy	74	Pro-----GlyAsnGluArgSerGlnGluLeuHis	83
Db	10450	CCGGCGGACATCCACCCCATCCCATCAGCGTGAACAACACCGACTCTCGACACACCCAC	10399
Oy	84	LeuArgProGlnSerHisSerTyrLeuProGlnLeuGlyIysSerGluMetGluPheIle	103
Db	10390	ATCAACCGACAAACACCAACCGACCGCGGCC-----	10361
Oy	104	GluSerIysArgProAlaGlyLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu	123
Db	10360	ATTCAACGGCGCGCAACCGAACCCATCTCC--CGAACCTCAACAACCCCGGACACCG	10300
Oy	124	LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIysAspArgSerLeuThrGly	143
Db	10301	ACTCCGACGCGCCCGACCGACCACTATCCCCACCCACCGGCAACCCACCATCAACCC--	10245
Oy	144	LysLeuGluProValSerProProSer-----ProPro	154
Db	10244	-----CACCCCGCGCACCAACCGCACCGCATCCCAACCAACACACCCCGGCA	10199
Oy	155	HisThrAsp-----ProGluLeuGluLeuValProPro--ArgLeuSerIysGln	170
Db	10190	CATACGCGACGACCACTCCCGCCACCGAATGACCAACACCAACGACATCTCCACACCC	10133
Oy	170	uGluLeuIleGlnAsnMetAspArgValAspArgGluIleIleThrMetValGluGlnGlnI	190
Db	10130	GAGCTC---CAACGCCCGGAAACA-----CGCACCTTCCAACGCAACACGAGGCT	10080
Oy	190	eSerIysLeuIysIysIysGlnGlnGlnLeuGluGluGlnAlaAlaIysPro-----	207
Db	10079	GAGCAACTCGTCCGACCCACCAACCAACCCCAAAACCGAGTCAAGT--CCGGAAATCT	10022
Oy	208	-----ProGluProGluIysProValSerPro-----	216
Db	10020	ACATCCGCGAACGACCGACCATCTCACTCCCGCCACGACGAAACACTCATCCAGACCC	9961
Oy	217	-----ProPro-IleGlu---SerIysHisArgSerIleValGlnIleI	230
Db	9960	TCCGCAACACCCACCCCGCGCTACCAACCGACCGCCCATCCACCAACCCACTCGGACCC	9901
Oy	230	IeTyrAspGluAsnArgIysIysIeGluIleAlaHisIeArgIleLeuGluGlyLeuGlyP	250
Db	9900	TGACCCCGGAACAACACCCCAACCCCAACCGACCCCGACCGC-----	9860
Oy	250	roGlnValGluLeuProLeuTyrIrsnGlnProSerAspThrArgGlnTyrHis-----	267
Db	9859	-----ACTCCCGCGACACACGACACCGGCCCCGGAAGACACACCCACAGAGACC	9811
Oy	268	-----GluAsnIleIysIleAsnGln-----AlaMetArgIysL	279
Db	9810	CCGGCAGCACCAATCTCAACCCCGCGACGAATCTCTCAAGCCCGCCCGCCACACACACC	9751
Oy	279	ysLeuIleLeuTyrPheIysArgIrsnHisIeAlaArgIysGlnTyrIrsnIysPheC	299
Db	9750	GCAGCTCATCGGAACGACG	9713

[illegible]

QY	659	argGlnaenleuabpGluIuleuGlnGlnIhIshlyseuIysemGluIyGluIyrgAsn	678
Db	9061	-----CTGATTGACCGCATGCGCCCGCACCCGCCA-GCACCCGAAGACCGTTGCGGC	9009
QY	679	AlaATgATgLyLyLyLyLyAlaProAlaAlaAsnGluGluAlaAlaPheProPro	698
Db	9008	GCGCGTCCG-----AAAGCCGCTTCACAGACAGACACCCACAC-----CTTCG	8964
QY	699	ValValGluabpGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetVal	718
Db	8963	CCCAAGCCGTCGCTCCG-----	8946
QY	719	GluGluAlaGluAlaLeuHlaAlaSerGlyAsnGluValProArgGlyGluCySerGly	738
Db	8945	-----CACC-CGCCGAACCGCTTGCACCCGCTCCGCGCGCAGC	8905
QY	729	ProAlaThrValaAsnSerSeraphrGluSerIle-ProSerPro-----	754
Db	8904	CCCCGCTCCCGAGAACTCCACGAACACCGTCGCGCTGCATACCTGACCCACCG	8845
QY	755	-----HisThrGluAlaAlaLyAspThrGlyGlnAsnGlyProLyProPr	770
Db	8844	GCAACCGCAGATCACTCAACCCCGCGCAACAGACTGACCGCAGATGCAAGCCAC	8785
QY	770	AlaThrLeuGlyAlaAspGlyProPro-----ProGlyProProThrPro-----	785
Db	8784	AACGACGAGAACCGCGGTGCCACCGTCAGCGCGCGCCCTCCAAACCCAAAGCATAG	8725
QY	785	-----	785
Db	8724	GCAACCGCAGCGAGATCACTGCGCGAGGTCCCGTGAGACATGCGCGCGTAGGTG	8665
QY	785	-----	785
Db	8664	TCCGCGCGGTCCCGCAGTGGGTGCGGTATCTCTGTGGAGATGCGCGTAAGACACCG	8605
QY	786	-----ProArgArgThrSerArgAlaProIleGluProThr-	797
Db	8604	GTCCGGGTGCGCGCAGCGAGTCCGGAACGATGCGCGCGCTCCAGCGCTCCACGAC	8545
QY	798	-----ProAlaSerGluAlaThrGlyAlaProThrPro-----	808
Db	8544	GTCCTCAGACCAACCGCTGCTGCGGGTCCATGAGCCGTGGCCTCGCGCGGATATCCG	8485
QY	808	-----	808
Db	8484	AAGAATCCCGCGTGAACACTCGCGCGCGGAACGCAAGAACCCGCTCCCGATCAATAGTC	8425
QY	809	-ProProAlaProProSerProSerAlaProProProValAlaProIyAsnGluIyGly	828
Db	8424	TTGCCCGGCAACCCCGGCTCCGGGTGTAAGCCCTCGATCTCCAGCCACGCGTGGCC	8365
QY	828	uGluGluThrAlaAlaAla-----ProProValGluGluGlyGluGluIyGly	844
Db	8364	GGAACGGGAGAGACGTGTGCTGCATCCGCGCACAGTCTCCACAGTCTCCGCGAC	8305
QY	844	sPro-----	845
Db	8304	GCCACACCGCGCGGTAAAGCAGCCCATGCCACAGATCGGATCGCTACCGCTCGTC	8245
QY	846	-----ProAlaAlaGluIuleuAlaValaAspThrGlyLyAsnGluIuProValIy	863
Db	8244	GCCATGACCGGCGTg-gggcgcgcgcgTgacggcgTcgcgTcgTgTgCGGAGCGAAG	8186
QY	863	sSerGluCyThrGluGluAlaGluGlu-GlyProAlaLyAsnGlyLyAspAlaGluAla	883
Db	8185	TTCTGTCTGCGAGTGGAGGAGCAGACAGAGGGGTCTCGGAGTCTCAAGACCA-----	8133
QY	883	IaGluAlaThrAlaGluIyAlaLeuLyAsnIaGluIyLyGluIyGlySerGlyIyrga	903
Db	8132	-----GCGTGTTCGCGAGGTTCAGCGCGCGTGCACGGCTTACCGAGTTCCAG	8078
QY	903	IaThrThrAlaLySerSerGlyAlaProIaAspSerAspSerSerAlaThrCySerA	923

[illegible]

Db	7114	-----ACGGCGGGCGCCGACCGGGGACCTGGC	7085
Qy	1205	GIyGlySerIleThrLySGlyIleProSerThraArgValProSerAap-SerAlaIleTh	1224
Db	7084	GGCGGACCTGCCACCGACGAGGC-----ACCGTGCGCCCGGGTGATCACTACGCTG--	7036
Qy	1224	rTYrArgIleSerIleThrHisGlyThrProAlaSerValLeuTyrlusysGlyThrIleTh	1244
Db	7035	-----CCGAGAGACACCGGTCGTCCGACAGACAC	7004
Qy	1244	ArgIleIleGly-----GluAerSerProSerxArgLeuAraArgIlyArgGluAerSe	1262
Db	7003	GTCCGAGGATGTCGTTCCTCGTGGCTCTCTCCACCGACGGCGGTGGC-GGGGGCTGTGTACGG	6945
Qy	1262	rLeuProLySGlyHisValIleTyrglyGlyLyLySGlyHisValLeuSerTyrglugl	1282
Db	6944	TGCGAGCGGAGGCGCG--GACGGGAGCGGGCGGAGAACACACGCGGACATGAGACGC	6888
Qy	1282	yGlyMetSerValThrGlnCySerLySGlyLeuAerGlyArgSerSerGlyProProHis	1302
Db	6887	CGGAGG-CGCCGACCG-----CGAATGATTTCTCC--	6855
Qy	1302	sglurhrrAlaIaProLyArgThrTyraSpmMetGluGlyAraGlyAlaI	1322
Db	6854	-CGCGAGACAGCGCCCGACAGACTACGCA-----GACCGCGGGCGAC	6811
Qy	1322	eSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerPr	1342
Db	6810	GCACCTC-----CGGCAAGTCCGCCAACCGCCCGACGCG	6775
Qy	1342	oHisHisLeuLySGlyGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSe	1362
Db	6774	CGCCCAAGTTCGAGGCGGCGGCACTGTCCGA-----GTCC--	6738
Qy	1362	rTyraGlnGlnGlnGluAerTyrluAraArgGlnAlaLyLeuLySerArgGluG	1382
Db	6737	-----ACACTGAAGCGCCCGACAGAAAGG	6712
Qy	1382	yThr-----	1383
Db	6711	GTCTGTGCGCGCGGACGCGCGCGCTCACGAGTGAGATTCAGAGCGGGGCTTCGGGT	6652
Qy	1384	-ProProProProProProSerArgAaP-	1392
Db	6651	ACCGCTCGCCGACGCGCTGACGAGCGGCGACGCGGTGCGCGCTCCGAGTCCGCGCGG	6592
Qy	1393	-----LeuThrGlu-----AlaTyrlusThrGlnAlaLeuGlyPr	1404
Db	6591	CGGAGCCACAGAACTCGGGGTGCTGCGGCGGCGGACCGTCAACCGCGCGGACAGCTTC	6532
Qy	1404	oLeuLySerLeuLyProAlaHisGluGlyLeuValAlaThraValyGlu-----	1420
Db	6531	C---CGGCTGCTCC--GCGCTGACCGACCGGACCGGACCGGTGTGAC	6484
Qy	1421	-----AlaGlyArgSerIleHisGluIleProArgGlnGluLeuAraGHisThrProG	1438
Db	6483	TGCGCGCCGCGGACGCGGAGTTCACTGCGATCGGGGGGCGAGTCCGGAAGTCCGCG--	6426
Qy	1438	uLeuProLeuAlaProArgProProLeuLySGlySerIleThrGlnGlyThrProLeuLy	1458
Db	6425	-----TGTCTCTGCTCC--GGGCGAAGACGAC	6400
Qy	1458	sTyraSPThrGlnGlyAlaSerThrThrGlySerLyLyHisAerValAArgSerIleIleG	1478
Db	6399	ACATACACGCGAGCGCCGGGGGTGGCCGCGGTGACGACGCGCGCTGCCAAGCAGCGCG	6340
Qy	1478	ySer-----ProGlyArgThrPheProProValHisProLeuAerValMetAlaAs	1495
Db	6339	TAGCGAGATCCCGCGGTGTGACCGTCCCGCTCAGTGGCGCCGAGCGTCCGCGCGAG	6280
Qy	1495	pAlaArgAlaLeuGlnGlyAlaCyetyrluGlySerLeuLySerArgGlyProGlyThrAl	1515
Db	6279	CGGGGGGTTCACGCGCG--TCACGCGGTGTCGACGAGGCTTCGAC	6235

QY	1515	ASerSerSerGlyGlySerIleAlaArgGly-	1525
DB	6234	CAGTAGC-----GGTCGGCTTGGAAAGGGTAGGTGGGAGGTGCATGCGCGGCGGTG	6181
QY	1526	-----AlaProValIleValProGluLeuG1	1534
DB	6180	CTCCCGGTGTGAGTAGTCCGTGGCTCCAGCGCGGCGTCCGGGTGAAACACCTGTGGCAGC	6121
QY	1534	YLysProArgGlnSerProLeuThrTYrGluAspHisGlyAlaProPheAlaGlyHisLe	1554
DB	6120	GCGCGCTTGAAACACTGCC-----GCTCGGACAGG	6091
QY	1554	uProArgGlySerProValThrMetArgGluProThrProArg-LeuGlnGluGlySerL	1574
DB	6090	CCCCCGCCGATCGCGGCGA-----CCACCAACCATCATCT-----	6056
QY	1574	euserSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu-----	1591
DB	6055	-----ACCGGCCCCAGGACACTCACCGGCATCCCGCATCCCGGTGACACCCCA	6013
QY	1592	-----lAlaLysSerProHisSerThrValProGluHisHisProHisProI	1608
DB	6012	TGCGGACCCACTCTCCACCAACGTCCCA-----ACCAGACCAAGCACACCCGCCACCCA	5956
QY	1608	lSerProTYrGlnHisLeuLeuArgGlyValSerGlyVal-----AspLeuT	1624
DB	5955	TCCGGAAACCCACCGCTCCCGGCGATGAGCAGCCCAATCCCGGATCCCGCACTCC	5896
QY	1624	YrArgSerHisGleProLeuAlaPheAspProThrSerIleProArg-----	1639
DB	5895	CCCGAACCACCCACCCGACCCGACACCCGACACCAACACCCGCGCGCACCCG	5837
QY	1640	-----GlyIleProLeuAspAlaAlaAlaTYrT	1650
DB	5836	ACCGAATCCCAACGACTCCACACCCCGGAACTCCCAACACCGCTCATCA-----	5781
QY	1650	YrLeuProArgHisLeuAlaProAsnProThrTYrPro-HisLeuTYrProProTYrLeu	1668
DB	5780	-----ACAGGATGATGAACCCATGCGACATTCACATCCACCCCGGCACTC	5738
QY	1670	-----IleArgGlyTYrProAspThrAlaAlaLeuGluAsnArgGlnThr	1684
DB	5737	CACCCCATCCCCCATCAGCGAGGCAAC-CCACCGACTCCAGA-----	5697
QY	1685	IleIleAsnAspTYrIleThrSerGlnIleMetHisHisAsnThrAlaThrAlaMetAla	1704
DB	5686	-----CACCAATATCACCGGACAAACACCAACCGACGCG	5666
QY	1705	GlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSerLeuAlaLeuAsnTYr	1724
DB	5663	GCCCATCTCA-----	5655
QY	1725	AlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuVal	1744
DB	5654	-----CCGCGCGAACCGACCCCACT	5634
QY	1745	ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTYrLeuProThrAla	1764
DB	5633	CCCCCAACCCCTCAACAAACCCCGCGACACCGACTCCGAGCGCCCAACGACATATCC	5574
QY	1765	ProGlnProPheSerSerArgHisIleSerSerSerProLeuSerProGlyGlyProThrHis	1784
DB	5573	CCCCACCAACGGGCAACCCACCATCAACCCGCGGCGACCA-----CAACCGCA	5520
QY	1785	LeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArg	1804
DB	5519	CCGCATCACCCCAACAGAACACCCCGCGACATACGACGACGACCACTCCCGCACCAAT	5466
QY	1805	AspArgAspArgGluArgGluGlySerIleLeuThrSerThrThrThrValGluHisAla	1824
DB	5459	GAC-----CCAAACACCAACGACACTCTCCACACCCC	5433

Db	3659	GGTGCTACAGCCGCTCAGAGTCCAG--CCGCGGTGCGTGGGGAAAGACAGATGGCGTCG	3601
Qy	2457	AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet	2476
Db	3600	GGC-----CCGAGCTCCAGACAGCTCCCGAGAGCCGCGCGC	3565
Qy	2477	AlaSerProPro-----ProGlyLeuProAlaGlySerGlyPro	2490
Db	3504	GAGTCGACGCCGCCCGGGAGCCGCGACGCCGACACCCACGATGCCATGCCGTCCGTCTCG	3505
Qy	2491	LeuAlaGlyProHisIleAlaTyrAsp-----GluGluProLysPro--	2504
Db	3504	TCCGCGCTGCGCCCGCGCGCTCCGACACCCCTACCGCGCGCCGACGAGACATCCGCGCG	3445
Qy	2505	-----LeuLeuGlySerGlnTyrGluThrLeuSerAspSerGlu	2517
Db	3444	CCGCGCGCTCGAAGCTTCTCTGCGCCGCTGTCGAGAGACGAGACCCGAGGGGAGG	3391

Search completed: November 12, 2005, 14:54:06
Job time : 41449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: November 11, 2005, 09:34:07 ; Search time 35711 Seconds
(without alignments)
3415.244 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQDVAGTWRATEPRYP.....WDEBPXPLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xdp
-Q=cg2_1/USFTO.spool_p/US09522753/runat_02112005.173640.20793/app_query.fasta.1.2695
-DB=GenBml -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCFMT=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USFR=US09522753 @CGN 1.1 14601 @runat_02112005.173640.20793 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBml.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13215	100.0	8561	9 AF113003	AF113003 Homo sapi
2	12978	98.2	8686	9 AF125672	AF125672 Homo sapi
3	12922	97.8	9053	6 AR447713	AR447713 Sequence
4	11067	83.7	8388	10 AF113001	AF113001 Mus muscu

5	10832.5	82.0	8544	10	AF125671	AF125671 Mus muscu
6	9894	74.9	7465	10	AF113002	AF113002 Mus muscu
7	9776.5	74.0	6339	6	C0722208	C0722208 Sequence
8	7885	59.7	5989	9	HSU37146	U37146 Human allen
9	4612	34.9	2930	9	S83390	S83390 T3 receptor
10	4194	31.7	7949	9	AB028970	AB028970 Homo sapi
11	4187	31.7	7780	10	MM035312	U35312 Mus musculus
12	4147.5	31.4	7940	6	C0847916	C0847916 Sequence
13	4147.5	31.4	7940	6	AX578069	AX578069 Sequence
14	4147.5	31.4	7940	9	AF044209	AF044209 Homo sapi
15	4043	30.6	8018	9	AF087856	AF087856 Homo sapi
16	4009.5	30.3	8959	5	AF495886	AF495886 Xenopus 1
17	3736.5	28.3	191100	10	AC139377	AC139377 Mus muscu
18	3479.5	26.3	254449	2	AC097560	AC097560 Rattus no
19	3445	26.1	218129	2	AC136560	AC136560 Rattus no
20	3344.5	25.3	219339	2	AC121005	AC121005 Rattus no
21	3178	24.0	2842	9	BC004326	BC004326 Homo sapi
22	3012	22.8	2964	10	BC047524	BC047524 Mus muscu
23	2910.5	22.0	6300	6	C0724013	C0724013 Sequence
24	2891.5	21.9	205283	9	AC073916	AC073916 Homo sapi
25	2187	16.5	3120	9	AF303586	AF303586 Homo sapi
26	2155	16.3	1937	5	BC054296	BC054296 Xenopus 1
27	2147	16.2	1896	5	BC082706	BC082706 Xenopus 1
28	2091.5	15.8	6328	10	AB093281	AB093281 Mus muscu
29	1900	14.4	6541	9	AB019524	AB019524 Homo sapi
30	1839	13.9	4285	10	MUSRIPI3G	L78294 Mus musculu
31	1687	12.8	1891	9	BC050594	BC050594 Homo sapi
32	1663.5	12.6	1850	9	BC058511	BC058511 Homo sapi
33	1645.5	12.5	1808	9	BC068996	BC068996 Homo sapi
34	1633	12.4	1741	9	BC056862	BC056862 Homo sapi
35	1628.5	12.3	3025	5	BC049302	BC049302 Danio rer
36	1611	12.2	2145	5	BC073704	BC073704 Xenopus 1
37	1530.5	11.6	1857	2	BC073253	BC073253 Xenopus 1
38	1396	10.6	161970	2	AC027706	AC027706 Homo sapi
39	1391	10.5	2914	6	C0850622	C0850622 Sequence
40	1391	10.5	2914	6	AK127788	AK127788 Homo sapi
41	1350.5	10.2	1152	5	AY498876	AY498876 Xenopus 1
42	1172.5	8.9	3997	9	AF303585	AF303585 Homo sapi
43	1162	8.8	133947	5	AL590153	AL590153 Zebrafish
44	1113	8.4	6650	6	AX677743	AX677743 Sequence
45	1099	8.3	11256	3	AF175223	AF175223 Drosophila

ALIGNMENTS

RESULT 1
AF113003
LOCUS
DEFINITION
Homo sapiens silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
AF113003
ACCESSION
AF113003.1 GI:4454551
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 8561)
Ordentlich,P., Downes,M., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M.
Unique forms of human and mouse nuclear receptor corepressor SMRT
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)

AND20946

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
SUBMITTED (11-DEC-1998) Gene Expression Laboratory, The Salk
Institute for Biological Studies, 10010 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
Source
Location/Qualifiers
1..8561
/Organism="Homo sapiens"

QY 381 GUGNGUUBNLEUGLULYSGLMETARGSLULEUALAVALIIEPROPMETLEUTYR 400
DB 1142 GAGCAGAGAACTCGTAGAAGCAATGCGCAGCTGGCGGTGATCCGCCATCTGTAC 1201
QY 401 ASPALAAPGNGINAGYIIELVYPHEIIAANMETANGIYLUUMETALAAPPMET 420
DB 1202 GACCTTACACGACGCCCATCAATCAATCAATGAACGGGCTTATGCGGACCCCATG 1261
QY 421 LYEVATLYRYSAPARGINVAIMEASMETTRPSERGUGLUGLULYSGIUTHRPHE 440
DB 1262 AAGGTGACAAAGCCCCAGCTCATGAACATGTGAGTAGAGAGAGAAACCTTC 1321
QY 441 ARGGLULYPHEMETGNIHISPROLYSAPNHEGLYLEUIEALASERPHLEUGIYARG 460
DB 1322 CGGAGAAAGTTACGACGACATCCCAAGAACTTGGCTGATCGCATCTCTGAGAGG 1381
QY 461 LYETHRALIAGLUCYEVALILEUTYRTYTRILEUTHRYLSAENGILUANTYLYS 480
DB 1382 AAGCAGTGGCTGAGTGGCTCTCTATTACTTACTGATGAAGAAATGAACTATPAG 1441
QY 481 SERLEUVALIARGASERTYRARGARGGLYLSERSEGINGLINGLINGIN 500
DB 1442 AGCTGTGAGACGAGACTATCGCGCGCGCAAGACCGACGCAACACGACGAG 1501
QY 501 GINGLINGLINGINGLINGINGLINGINPROMETPROARGSERSEGINGLULYS 520
DB 1502 CACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1561
QY 521 ASRGGLULYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULYSG 540
DB 1562 GATGAG 1621
QY 541 ASPLEGLUAPLEULEULYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULY 560
DB 1622 GACAGAGAAAGCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 LYEGIULALAVAIASERLYSGIYARGLYETHRALIANSERGINGLYARGARGLY 580
DB 1682 AAGAGAGCTGTGGCTCCAAAGGCCGCAAACTGCCAAGCCAGGAGAGAGAGAG 1741
QY 581 ARGIIETHRYRGERMETALAAENGILUALAANSEGLUGIULALIEETHRPROGLIN 600
DB 1742 CGCATCAACCGCTCAATGAGTGAAGGCAACGACGAGGCGCATACCCCCACAG 1801
QY 601 SERALGLULEUALASERMETGULEUANGLUSERSERATRTPTHRGUGLUMET 620
DB 1802 AGCGCCAGCTGGCTCATGAGACTGAATGAGATTCGTCTGAGCAAGAGAAATG 1861
QY 621 GLUTHRALALYBLYSGIULEUGLUIHISGLYARGANTRPSERALIIEALARGMET 640
DB 1862 GAAGAGCCCAAGAAAGGTCTCTTGGAACAGCGCGCAACTGTGCGCATCGCCGATG 1921
QY 641 VALIGLYSERLYSTRVALISERGINCYELYSAPNPHETYPHEANTYRILYBLYARGIN 660
DB 1922 GTGGGCTCCCAAGACTGTGTGCGAGTGAAGAACTTCACTTCAACTCAAGAGAGCAG 1981
QY 661 ASPLLEUASPGIULILEUGINGIINHISLYLEULEULEWETGULYSGIULYSGIULY 680
DB 1982 AACCTGATGAGATCTTGACAGACCAAGCTGAAGTGAAGAGAGAGAGAGAGAGAG 2041
QY 681 ARGLYELYBLYBLYBALPROALAAIALAISERGLUGIULALAPHEPROPROVAL 700
DB 2042 AGGAGAGAGAGAAAGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 2101
QY 701 GLUAPRGUGIUMETGULIASERGLYVALISERGLYANGLUGIUGIUMETVALIGLU 720
DB 2102 GAGAGATGAGAGATGAGAGCGTGGCGGTGAGCGGAATGAGAGAGAGAGAGAGAGAG 2161
QY 721 ALAGIULALAEUHSBALASERGLYANGLULVALIPROARGIYGLIUCYSESERGLYPROALA 740
DB 2162 GCTGAAGCCTTACATGCTCTGTGGAAATGAGGTGCCAGAGGGAGATGACAGAGGCC 2221
QY 741 THRVALASNAANSESERAPTHRGLUSERILEPROSEPROHISTHRGLUALAIALYS 760

DB 2222 ACTGTCAACAAACAGCTCAGACACGAGAGCATCCCTCTCTCACATGAGGCCGCAAG 2281
QY 761 ASPTHRGLVGINANGIYPROLYSAPROVALATHRLEUGIYALIAAPGLYPROPRO 780
DB 2282 GACACAGAGGAGAAATGGAGCCCAAGCCCCAGCCACCTCTGGAGCGGAGAGGCCACCCCA 2341
QY 781 GLYPROPROTHRPROVALARGATHRSEARGLAIPROILEGIUPROTHRPROALASER 800
DB 2342 GGCCCAACCAACCCACACGAGAGACATCCCGGGCCCCCATTTAGGCCACCCCGGCTCT 2401
QY 801 GLUALATHRGLYALAPROTHRPROPROPROVALAPROSEPROSERVALAPROPRO 820
DB 2402 GAAGCCACCGGAGGCGCTTACGCCCCACAGACCCCATGCGCTCTGCACTCTCTCT 2461
QY 821 VALVALPROLYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULY 840
DB 2462 GTGGTCCCAAG 2521
QY 841 GUGLUGIULYSPROPROVALIAGLUGIULEUALAVALIAPTHRGLYLYSALAGLU 860
DB 2522 GAGGAGCAGAAAGCCCCCGCGCTGAGAGACTGGCATGACACAGAGAGCCGAGAGAG 2581
QY 861 PROVALYSESGIUCYSTRHGLUGIUALIAGLUGIULYPROVALYSGIULYSAAPALA 880
DB 2582 CCGGTCAAGAGCGAGTGCAGAGAGAGCCGAGAGAGGCGCGGCAAGGAGAGAGAGAG 2641
QY 881 GLUALIAGLUALATHRALIAGLUYALILEULYBALIAGLUYBLYSGIULYGLYSER 900
DB 2642 GAGGCGCTGAGGCAAGGCGCGAGAGGCGCTCAAGGCAAGAGAGAGAGAGAGAGAG 2701
QY 901 GLYARGIATHRTHRALALYSERSESGIYALIPROGINAPSEARAPSEARSEAR 920
DB 2702 GCGAGGGCCACACTGCAAGAGCTCGAGGCGCCCCCGAGACAGCGACTCAAGTCTAAC 2761
QY 921 CYSESERIALAAPGLUVALIAPGLUVALIAGLUYGLYLYSAPLYSAPNARGLEULEUSER 940
DB 2762 TGCAGTGCACAGAGAGTGAATGAGGCGGAGGCGGAGCAAGAAACGGCTGTCTCCCA 2821
QY 941 ARGPROSELEULEUTHRPROTHRGLYASAPPROVALIANSAPLEASERPROGLINLYS 960
DB 2822 AGGCCAGCTCTTCAACCCGAGCTGGAGACCCCGGCGCAATGCTTCAACCCCAAGAACCA 2881
QY 961 LEUASPLEULYSGIULEULYSGIULYSGIULYSGIULYSGIULYSGIULYSGIULY 980
DB 2882 CTGAGCTGAAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2941
QY 981 VALIAGLUPROPROARGIUALAPALAPROTHRLYSAPROVALAPROVALAPROPRO 1000
DB 2942 GTCATGAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3001
QY 1001 PROPROGLINANLEUGINPROGLUSERASAPALAPROGLINPROGLYSESERPROARG 1020
DB 3002 CCAACCGCAAACTTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
QY 1021 GLYLYSESERARGSERPROVALAPROVALAPROVALAPROVALAPROVALAPROVAL 1040
DB 3062 GGCAG 3121
QY 1041 LYLEUUPROGLYASAPROPROCYSTRPTHRSEGLYLEUPROPHAPROVALAPROPROARG 1060
DB 3122 AAGCTGCTGGGAGACCCCTTGTGATCTCGAGCTGCTGCTTCCCGGAGAGAGAGAG 3181
QY 1061 GLUVALIILEYVALIASERPROHISALAPROASAPROSEARALAPHESESTRYALAPROPRO 1080
DB 3182 GAGGTATCAAGAGCTTCCCGCATGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3241
QY 1081 GLYHISPROLEUPROLEUGIYLEUHSAPSTRALARGPROVALIENUPROARGPROPRO 1100
DB 3242 GGTACACCACTGCGCTGGGCTCTCATGACACTGCGCGGCGCTCTGCGCGGCCACCC 3301
QY 1101 THRILESERANPROPROLEULIASESERALYLYHISAPROSEARVALLEUGIYARG 1120

Db 3302 ACCATCTCCAAACCCGCTCCCTCATCTCTGTCCAGACACCCGAGCTCTCGAGAG 3361
Qy 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlnHis 1140
Db 3362 CAATATAGTGGCATCTCCCAAGGATGTGGTCCAGCTCCAGTCCCCGTAATCTAGAGCAT 3421
Qy 1141 AlaValAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProIleVal 1160
Db 3422 GCCAAGGCCCGGTGGGCTCTGTCAACATGGGGCTCCCTGGCCATGAGACCCCAAAAG 3481
Qy 1161 LeuAlaProPheSerGlyValIleGlnGlnGlnIleLeuSerProArgGlyValAlaGlyPro 1180
Db 3482 CTGGACACCTTCAACGCGAGTGAAGCAGAGCACTCTCCCAAGCGGCAAGCTGGAGCA 3541
Qy 1181 ProGlnSerLeuGlyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeu 1200
Db 3542 CCGAGAGCTCTGGGGGTGCCACAGCCCAAGAGGCTCCGTGTGAGAGGAGACGCTCTG 3601
Qy 1201 GlySerValProGlyValSerIleThrIleValIleProSerThrArgValProSerAsp 1220
Db 3602 GGCTCAGTTCGGGGCGGAAGCATCACAAAGGCAATCCCAAGCACAGGGTGCCTCGAC 3661
Qy 1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrVal 1240
Db 3662 ACGGCATACATACCGCGCTTCATACCCAGCAAGCGCACTGATCTCTGAACAAG 3721
Qy 1241 GlyThrIleThrArgIleIleGlyIleAspSerProSerArgLeuAspArgIleArgGly 1260
Db 3722 GGCACCATACCAAGATCATCGCGAGAGACGCCCAAGTCTTGAACCGCGGCGGGAG 3781
Qy 1261 AspSerLeuProIleGlyHisValIleTyrGlnGlyValGlyValIleLeuSerTyr 1280
Db 3782 GACAGCTGCCCAAGGGCCAGCATCTACGAAGGCAAGGCGCAAGCTTGTCTAT 3841
Qy 1281 GlnGlyGlyMetSerValThrGlnCysSerValAspGlyValArgSerSerGlyPro 1300
Db 3842 GAGGAGTGAATGTCTGTGACCGTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAG 3901
Qy 1301 ProHisGlyThrAlaAlaProIleAspGlyThrTyrAspMetMetGlnGlyValArgVal 1320
Db 3902 CCCCATGAGACGGGCCCCCAAGGCCACTATGATGATGAGGGCGCGTGGGCAAG 3961
Qy 1321 AlaIleSerSerAlaSerIleGlnGlyLeuMetGlyValArgAlaIleProProGlnArgHis 1340
Db 3962 GCATCTCTCTCAAGCCAGCATCGAAGGTCTCATGGCGGTGCATCCCGCGAGCCAGAC 4021
Qy 1341 SerProHisHisIleuIleGlnGlnHisIleHisIleArgGlySerIleThrGlnGlyIlePro 1360
Db 4022 AGCCCCACCACTCAAAAGCAGACCAATCCGCGGTCCATCACACAGGGATCCCT 4081
Qy 1361 ArgSerTyrValGlnAlaGlnGlnAspTyrLeuArgArgGlnAlaIleHisLeuLeuValArg 1380
Db 4082 CGGTCTTAAGTGAAGGACAGAGAGACTACTCGTCCGGAGGCCAAAGCTCTCTAAAGCGG 4141
Qy 1381 GlnGlyThrProProProProProProSerArgAspLeuThrGlnAlaTyrIleThrGln 1400
Db 4142 GAGGGACAGCTCCGCCCAAGCGCTCTCAAGGACTGACCGAGCTCTCAAGAGAGAGAG 4201
Qy 1401 AlaIleGlyProLeuIleValSerProAlaHisGlnGlyLeuValAlaThrValIleValGln 1420
Db 4202 GCGCTGGGCGCTGAGAGCGGCGCCATGAGAGGCTGTGGTGCACAGGAGAGAGAG 4261
Qy 1421 AlaGlyArgSerIleHisGlnIleProArgGlnGlyIleLeuArgHisIleThrProGlnLeuPro 1440
Db 4262 GCGGGCGCTCTCATGATGATCCGCGGAGAGACTGGGACACAGCCCGAGCTGCC 4321
Qy 1441 LeuAlaProArgProLeuIleGlnGlySerIleThrGlnGlnIleThrProLeuIleTyrAsp 1460
Db 4322 CTGGCCCGCGGCGCTCAAGAGAGGCTCATACAGAGGAGACCCCGCTCAAGTACAGAC 4381
Qy 1461 ThrGlyAlaSerThrThrGlySerIleValHisIleAspValArgSerLeuIleGlySerPro 1480
Db 4382 ACCGGCGGTCCACCACTGGCTCAAAAGCAAGAGTACGTCCTCATCGGAGAGCC 4441

Qy 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGln 1500
Db 4442 GCGCGAGAGCTTCCACCCGTGACCCGCTGATGTATGAGCCAGACGCCCGGACACTGGAA 4501
Qy 1501 ArgAlaCysTyrGlnGlnSerLeuIleValSerArgProGlyThrAlaSerSerSerGlyVal 1520
Db 4502 CGTGCCTGTACAGAGAGAGCTGAAGAGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAG 4561
Qy 1521 SerIleAlaArgGlyAlaProValIleValProGlnLeuGlyValSerProArgGlnSerPro 1540
Db 4562 TCCATTGGCGCGGCGCGCCCGGTCTTGCTGAGAGCTGGTGAAGCGCGGAGAGAGAGAG 4621
Qy 1541 LeuThrTyrGlnAspHisGlyAlaProPheAlaGlyHisIleLeuProArgGlySerProVal 1560
Db 4622 CTGACCTATGAGAGACACAGGGAGACCTTGTGCGGAGCACTCCACAGAGTTCGCGCGT 4681
Qy 1561 ThrMetArgGlnProThrProArgLeuGlnGlnGlySerLeuSerSerSerIleValAsp 1580
Db 4682 ACCATGGCGGAGCCCAAGCGCGCTGAGAGAGGCGAGCTTTCGTCCAGCAAGGATCC 4741
Qy 1581 GlnAspArgIleLeuThrSerThrProArgGlnIleAlaIleSerProHisSerThrVal 1600
Db 4742 CAGGACGGAAGCTGACGTGACAGCTCTGATGATGCCAATCCCGCACAGACCGT 4801
Qy 1601 ProGlnHisIleProHisProIleSerProTyrGlnHisIleLeuLeuArgGlyValSerGly 1620
Db 4802 CCGAGACACACCCACACCCATCTCGCTTGAAGCACTGCTTGGGGCGTGAAGTGC 4861
Qy 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 4862 GTGAGACTGTATGAGAGCCATACCCCTGGGCTTTCAGACCCCACTTCATACCCGCGG 4921
Qy 1641 IleProLeuAspAlaAlaAlaIleTyrIleuProArgHisIleuAlaProAspProThr 1660
Db 4922 ATCCCTGTGAGAGAGCGCTGCTATACCTGAGCCCGCACACTGGCCCCCAACCCAC 4981
Qy 1661 TyrProHisIleuTyrProProTyrIleuIleArgGlyTyrProAspThrAlaAlaLeuGln 1680
Db 4982 TACCGCACCTGTACCCACCTTACCTCATCCGCGGTACCCGAGACAGCGCGGTGAG 5041
Qy 1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnIleMetHisHisAsnThrAla 1700
Db 5042 AACCGGACACCATCATCATATATACATCACTGACAGAGATGACACACACAGAGAGAG 5101
Qy 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSerLeu 1720
Db 5102 ACCGCATGGCCAGAGAGCTGATATGCTGAGGGGCTCTCGCGCGGAGTCTCGCTG 5161
Qy 1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisIleu 1740
Db 5162 GCACTCACTAGCTGCGGGTCCCGAGCATCATGACCTGTCCCAAGTGCACACCTG 5221
Qy 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
Db 5222 CTGTGCTGTGCTGCGCCGACACAGGACCCCAAGCCCAAGGAGAGAGAGAGAGAGAG 5281
Qy 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db 5282 CTCCCAACCGCGCCAGAGCTTTCAGACAGCCGCAAGAGCTCCCACTCTCCAGGA 5341
Qy 1781 GlyProThrHisIleuThrIleValProThrThrThrSerSerSerGlnArgIleArgAspArg 1800
Db 5342 GGTCCAAACACTTGCAAAACCAACACACAGTCTCTCCGAGCGGAGAGAGAGAGAG 5401
Qy 1801 AspArgGlnArgAspArgAspArgGlnArgGlnValSerIleLeuThrSerThrThrThr 1820
Db 5402 GATCGAGAGCGGAGCGGAGTGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5461
Qy 1821 ValGlnHisAlaProIleTyrArgProGlyThrGlnGlnSerSerSerGlySerSerGlySer 1840
Db 5462 GTGAGACAGACCCATCTGAGAGACTGTATCAGAGCAAGAGAGAGAGAGAGAGAGAG 5521

QY 1841 SerGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5522 AGCGCGCGGGGTGGGGGACAGACGCGCCCGCTCCACTCCATGCCACACGAC 5581
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5582 TCGGCCACTCTCCCTCGGACCCAGATGCTCTCCAGCAGAGACCAGTGTCTTACACAC 5641
QY 1881 ThrGlyMetIysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSer 1900
DB 5642 ACAGAGCATGAAGGGTATCATACCGCTGTGAGCCGACAGACCCACGCTCTTAGTCC 5701
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5702 ACCCTCCACTCTCTCACTCCGCTGCGCCAGCTGCACATTCCTCCAGTCCACCTGCCC 5761
QY 1921 LeuGlyGlyThrLeuAspGlyValTyrProThrIleuMetGluProValLeuLeuProLys 1940
DB 5762 CTGGGCGGACCTCTGATGGGGTGTACCTCACTGAGAGCCGCTTGTGCTCCAG 5821
QY 1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
DB 5822 GAGGCGCGCGGGTGGCGCGGACGAGCGCCGACAGACACCGGCATGCTCTTCTC 5881
QY 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
DB 5882 GCCAAGCGCCGAGCGCTCGGGCTGAGCGCGCTCTCTCCACAGAGGCTCGGAG 5941
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaGlnThrProAlaLys 2000
DB 5942 CCGCGCGCTTATGCTCTCTGTCGCGACCCACACATCGCCGACCCCTCGCAG 6001
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerLysAsp 2020
DB 6002 AACCTGCACCTACACACGCGACGCGCGGACCGCGCGCTGCGCTCGGCTCGGAC 6061
QY 2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuArgSer 2040
DB 6062 CCGCACCGGAGAAAGACTCAAGTAACCTTTTCCATCCAGGACTCGAACTCGTTCT 6121
QY 2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
DB 6122 CTGGGTTACACGCGACGACACTACGCGCGAAGGGGTGAGCGCGCTGACCTGTAGC 6181
QY 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisIleuGlnGluLeuAspLysSer 2080
DB 6182 TCACCCAGCTTGACCCGACGACAGGGGCTCCCAAGCACTGGAAAGCTTCGACAGAGC 6241
QY 2081 HisLeuGlnGlyLeuLeuArgProLysGlnProGlyProValLysLeuGlyGlyAla 2100
DB 6242 CACTGGAGGGGAGCTGGCGCCCAAGCAGCAGCGCCCGTGAAGCTTGGCGGGAGGCC 6301
QY 2101 AlaHisLeuProHisIleuLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
DB 6302 GCCCAGCTCCACACCTGCGCGCTGCTGAGAGCAGCGCTCTGTCACCGCTGCTC 6361
QY 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
DB 6362 CAACCGCGCCCAAGGGGTCAAGGTCAACAGCGGTGTCACTTGGCGCCAGCATAGT 6421
QY 2141 GluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro 2160
DB 6422 GAGGTATCACACAGGACTACCGGCAACCCAGCAGAGGTGAGCGGACCCCTGCCC 6481
QY 2161 AlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgProPro 2180
DB 6482 GCCCGCTTACTCTCTCTGAGGCGAGTGCCTCTCTGAGACCTCCCGCGCCACCC 6541
QY 2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgLysSerProHisSer 2200
DB 6542 AGTGACCTTACTCTCCGCGCGGACGACATGTGCCCCGCGCTGCTCCCCACAGC 6601
QY 2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220

DB 6602 GAAGGGGCAAGAGGTCTCCAGAGCCAAAGACGTGCTTGGGGTGGTGAAGAC 6661
QY 2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
DB 6662 GGTATTAACCTGTGTCCACCGGAGGGCATGACGAGCCAGGGCACTCCGGAGTGTCT 6721
QY 2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys 2260
DB 6722 GTTATCCCGCTGCTGTACCGGATGGGGACAGCGAGCCACAGATGGCTCCAG 6781
QY 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
DB 6782 TCTCCAGGCAACACCAACGCGCGGACGCTTCTTTCAGAACTGACGAGAGCAACTCC 6841
QY 2281 AlaMetValLysSerLysGlnGlnIleAsnLysIleuAsnThrHisAsnArgAsn 2300
DB 6842 GCCATGGTCAAGTCAAGAAAGCAAGATCAACAAAGAGCTGAACCCACACCGGANT 6901
QY 2301 GluProGluTyrAsnIleSerGlnProGlyTyrGluIlePheAsnMetProAlaIleThr 2320
DB 6902 GAGCTGAATACATATACGACGCTGGAGCGAGATCTTCAATATGCCCGCATAC 6961
QY 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMet 2340
DB 6962 GGAACAGGCTTATGACTTATGAGACGACGCGGTGAGGAAATGCCACACCAACATG 7021
QY 2341 GlyLeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrThrAspGlnTyrGluLys 2360
DB 7022 GGGCTGAGGCGCAATTAATTAAGGACCTCATGGTAAATATGACCTGGGAAAGTCTC 7081
QY 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLysLeuProAlaAla 2380
DB 7082 CCGCGCTCAGCGCCCAATGCTTTTAACTCTTAATCCATGCTGCGAGCTGCGCTGCT 7141
QY 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
DB 7142 ATGCCATACCGCTGCTGACGAGGAGTACACACACTCACTGCGCAGGTGCGGC 7201
QY 2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly 2420
DB 7202 GGGAAAGGCCAAGGTCTCTGGCAGACCCAGCAGCGAAACCAAGTCCCGCGCCCGG 7261
QY 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlnGlyAspCysAsn 2440
DB 7262 CTGGCATCTGGGACCGGCCACCTCTGTCTCTCACTGATGACTCGAGGAGACTGCAAC 7321
QY 2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
DB 7322 CGCGGAGCGCGCTCAACACCGGTGGGAGAGAGGCGCTGTCGCGAGGTTCCAGC 7381
QY 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
DB 7382 CCAATCCCTTACAAACCCCTGATCATGCGCTGACGCGGGTGTCAATGCTTCCACCC 7441
QY 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTyrAspGlu 2500
DB 7442 CCACCGGCTCTCCCGCGGAGCGGCGCTGCTGCGGCCCAACAGCTTGGAGAG 7501
QY 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLysSerAspSerGlu 2517
DB 7502 GAGCCCAAGCACTGCTGCTGAGTACGAGACACTCCGACAGCGAG 7552

RESULT 2
AF125672 8686 bp mRNA 1linear PRI 04-APR-1999
LOCUS AF125672
DEFINITION Homo sapiens silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform (SMRTS) mRNA, complete cds.
ACCESSION AF125672
VERSION AF125672.1 GI:4559297
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

AAD 22973

Db 997 ATCTTGACTTCAAGAGGAGAAATCACGCTCGAAGAAACAATGGAGACAAAGTCTTCGACG 1056
Qy ArgTyrAspGlnLeuMetGluAlaLeuGluTyrValGluIleArgTyrGlnLeuAsnAspPro 320
Db CGCTTATGACCAAGCTCTGAGAGGCTTGGAGAAAGAGTGGAGCCGATCTGAGAAACAACCC 1116
Qy ArgTyrArgAlaValysGlnSerLysValArgTyrTyrTyrGluTyrGlnPheProGluIle 340
Db CGGCGCGGGCCAGAGAGACAAAGTTCCGAGTACTGAGAAAGCAAGTTCCTGAGATC 1176
Qy ArgTyrGlnIleArgTyrGlnLeuGlnIleArgMetGlnSerArgValGluGlnIleArgTyrSerGly 360
Db CGCAAGAGCGCGGCTGACAGAGCGCATGCGAG---AGGGTGGCGCACGCGGAGCATGCGG 1233
Qy LeuSerMetSerAlaAlaArgSerGluHisGlnValSerGluIleIleAspGlyLeuSer 380
Db CTGTCCATGTGCGCCGCCCGAGAGACAAAGTGTGACAGATCATGATGAGCTCTCA 1234
Qy GlnGlnGlnIleAsnLeuGlnIleArgMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
Db GAGCAGAGAACTGAGAAAGACATGCGCCAGCTGGCCGTGGATCCGCGCCATCTGTAC 1294
Qy AspAlaAspGlnGlnIleArgTyrIleValPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
Db GACGCTACCAAGCGCCGATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1354
Qy LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnIleValysGluTyrPhe 440
Db AAGGTGTACAAAGCCCGCAGGTCAATGAATGTGATGAGAGAGAGAGAAAGACCTTC 1414
Qy ArgGluTyrPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db CGGAGAAAGTTACAGACGACATCCCAAGAACTTTGGCTGTAGTCCGATCTCTGAGAGG 1474
Qy LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db AACAACGTGCTGAGTCCGTCTCTATTACTGACTGACTGAAGAAATGAGAACTATAG 1534
Qy SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db AGCTGTGTGAGAGAGACTATCGCGCGCGAGAGAGCCAGCAGCAACACAGCAGCAG 1594
Qy GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db CACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1654
Qy AspGluTyrGlnIleArgTyrGlnIleArgTyrGlnIleArgTyrGlnIleArgTyrGlnIle 540
Db GATGAGAAAGAGAAAGAAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1714
Qy AspLysGlnAspLeuLeuLeuLysGluTyrThrAspAspThrSerGlyValAspAsnAspGlu 560
Db GACAGAGAAACCTCTCAAGAGAAAGACAGACCTCAGGGAGAGCAACGAGAG 1774
Qy LysGluValAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleArgTyrGly 580
Db AAGAGAGCTGTGGCTTCAAGCGCCGCAAACTGCCAAGCAGCAGGAAACCGCAAGGC 1834
Qy ArgTyrThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 600
Db CGCATCAACCCGCTCAATGAGCTAAATGAGGCAACGAGAGAGGCAATCAACCCCGAG 1894
Qy SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlnGluGluMet 620
Db AGCGCGAGCTGGCTTCCATGAGCTGAAATGAGAGTTCTGCTGAGAGAGAAAGAAATG 1954
Qy GlnThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
Db GAAAGAGCCAAAGAAAGGTCTCTGAGAACAGCGCGGCACTGGTGGCACTGCGCGAGT 2014
Qy ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysValArgGln 660
Db GTGGGCTCCAAAGCTGTGTGCGAGTGAAGAACTTCTTAACCTCAAGAAAGAGCAG 2133

Qy AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluTyrGluArgAsnAlaArg 680
Db AACCTCATGATGATCTTTCAGACAGACAGCTGAAGATGAGAAAGAGAGAGAGAGAG 2134
Qy ArgLysLysLysValAlaProAlaAlaAspSerGluValAlaPheProValVal 700
Db AGGAAAGAAAGAAAGAGCGCGCGCGCGAGAGAGAGGTGATTCGCCCGGTGGT 2194
Qy GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db GAGGATAGAGAGATGAGAGCGGTGCGGCTGACCGGAAATGAGAGAGATGTGAGAGAG 2254
Qy AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlyCysSerGlyProAla 740
Db GCTGA-----GCC 2314
Qy ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisThrGluAlaLys 760
Db ACTGTCAACAAAGCTCAGACACCGAGAGCATCCCTCTCTCACTGAGGCGCGCAAG 2323
Qy AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGluValAspGlyProProPro 780
Db GACACAGGACAGAAATGGGCCCAAGCCCGACCACTGGGCGCGGACCGGCGCA 2442
Qy GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db GGGCCACCCACCCCAACCGAGAGACATCCCGCC-CCCACTGATCTCACCCGCGCTC 2443
Qy ArgAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db TGAAGCACCTTAGCCCTTAGCCCTAGCCCGCACAGACCCCATTTCCCTTCACTCC 2502
Qy OvalValProLysGluGluTyrGluGluGluThrAlaAlaAlaProProValGluGlu 840
Db TGTGTCTCCCAAG 2562
Qy YGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 860
Db GAGAGACAAAG 2622
Qy UpProValysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db GCCGTCAAG 2682
Qy ArgAlaAlaAlaGluAlaThrAlaGluGluAlaLeuLysAlaGluLysGluGlySe 900
Db GGAAGCGCTGAG 2742
Qy ArgLysArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Db CGGAGAGGACCAACAGCCAAAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2802
Qy rCysSerAlaAspGluValAlaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 940
Db CTGCAGTGCAG 2862
Qy OArgProSerLeuLeuThrProThrGlyAspProAlaAlaAsnAlaSerProGluLysPr 960
Db AAGGCCAGCTCTCTCAACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2922
Qy OleuAspLeuLysGluLeuLysGluAlaAlaAlaIleProProIleGlnValThrLys 980
Db ACTGACCTGAGACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3041
Qy ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Db AGTCCATGAG 3042
Qy OProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
Db GCCACGCAAAACCTGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3102

[illegible]

QY 2472 nAAGlYValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
| | | | |
Db 7542 GGGGGGTGTCAATGGCTTCCCAACCCCAACCGGGGCTCCCGGGGGAGCGGCCCTGCGC 7601
QY 2492 AGlyProHisHisAlaATPAPAGlUGlUProlySPProLeuLeuCySerGlnTyGluTh 2512
| | | | |
Db 7602 TGGCCCCCAACCGCTGGGAGAGAGAGCCCAAGCACAATGCTGTCTGCAAGTACAGAC 7661
QY 2512 rLeuSerAPSerGlu 2517
| | | | |
Db 7662 ACTCTCCAGACGAG 7677

RESULT 3
AR447713 9053 bp DNA linear PAT 20-FEB-2004
LOCUS Sequence 306 from patent US 6673549.
DEFINITION AR447713
ACCESSION AR447713 GI:42676037
VERSION AR447713.1 GI:42676037
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 9053)
AUTHORS Furness, L.M. and Buchbinder, J.L.
TITLE Genes expressed in C3A liver cell cultures treated with steroids
JOURNAL Patent: US 6673549-A 306 06-JAN-2004;
FEATURES Location/Qualifiers
source 1. 9053
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:

Pred. No.: 1 09e-252 Length: 9053
Score: 12922.00 Matches: 2487
Percent Similarity: 98.26% Conservative: 2
Best Local Similarity: 98.18% Mismatches: 10
Query Match: 97.78% Indels: 35
DB: 6 Gaps: 5

US-09-522-753-5 (1-2517) x AR447713 (1-9053)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyPro 20
| | | | |
Db 520 ATGTGGGATCCACACAGCCTGTGGCAGACGTGAGGGCCACTGAGCCCGCTACCCG 579
QY 21 ProHisSerLeuSerTyProValGlnIleAlaTyThrHisThrAspValGlyLeuLeu 40
| | | | |
Db 580 CCCCAAGCCTTCTTACCCAGTGCAGATCGCCCGACACACGACGCTCGGGCTCTTG 639
QY 41 GluTyrglnHisHisSerArgAspTyralaSerHisLeuSerProGlySerIleIleGln 60
| | | | |
Db 640 GAGTACACACACTCCCGGACATATGCTCCCACTGTCCCGGCTCCATCATCCAG 699
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
| | | | |
Db 700 CCCAGCGCGGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGAAATGAAGCTCCAG 759
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrlleuProGlyLeuGlyLysSerGluMet 100
| | | | |
Db 760 GAGCTCCACCTGGCGGCAAGTCCCACTACATCTCCGAGCTGGGAAATCAGAGATG 819
QY 101 GluPheIleGlySerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
| | | | |
Db 820 GAGTTTATTGAAAGCAAGGCGCTCGGCTHAGCTGCTGCTAACCCTGCTGCCAGCG 879
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
| | | | |
Db 880 TCAACCCCTGCTGGCGCAGGGCCAGCTGGGGATCGAAGACTCAACCAAGACCGTAGC 939
QY 141 LeuThrGlyLysLeuGluProValaSerProProSerProProHisThrAspProGluLeu 160
| | | | |
Db 940 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCCAGCCCGCCGACACTGAGCTTG 999

QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
| | | | |
Db 1000 GAGCTGTGTCCCGCCAGCGGCTGTCCAAGGAGGAGCTGATCCAGAAACATGGAACCGCTGAGC 1059
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
| | | | |
Db 1060 CGAGAGATCAACCATGTGTAGAGAGAGATCTTAACTGTAAAGAAAGAGAGAGAGAGCTG 1119
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValaSerProProIleGlu 220
| | | | |
Db 1120 GAGGAGAGGCTGCCAAGCCCGCCAGCTGTGAAGAGCCGTGTACCGCCGCTCATCGAG 1179
QY 221 SerLysHisArgSerLeuValGlnIleIleTyraSPGluAsnArgLysAlaGluAla 240
| | | | |
Db 1180 TCGAAGACACCGCAGCCTGGTGAGATCATCTACAGCAGAAACCGAAGAGCTGAAGCT 1239
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrrAsnGlnPro 260
| | | | |
Db 1240 GCACATTCGGAATCTGGAAGGCTGGGGCCCGAGGTGAAGCTGCCCTGTACCAACCGCCC 1299
QY 261 SerAspThrArgGlnTyrrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
| | | | |
Db 1300 TCCGACACCCCGCACTATCATAGAAACATCAAAATTAACACAGCGATGCGAAGAGCTA 1359
QY 281 IleLeuTyrrPheLysArgArgHisHisAlaArgLysGlnTyrrLysGlnLysPheCysGln 300
| | | | |
Db 1360 ATCTTGTACTTCAAGAGAGAGAAATCACCGCTCGAAACAATGGAGACAGAAATGTTGCCAG 1419
QY 301 ArgTyrrAspGlnLeuMetGlnAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
| | | | |
Db 1420 CGCTATGACACAGCTCATGAGAGGCTGGGAGAAAGTGAAGCGCATTCAGAAACACCCC 1479
QY 321 ArgArgArgAlaLysGlySerLysValArgGluTyrrTyrgluLysGlnPheProGluIle 340
| | | | |
Db 1480 CCGCGGCGGGCCAAAGAGAGCAAGGTGCCGAGTCTACAGAAAGAGATTCCCTGAGATC 1539
QY 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
| | | | |
Db 1540 CCNAGACAGCGCAGCTGAGAGCGCATGACAGACAGGCTGGCGCAGCGGGCACTGTGG 1599
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValaSerGluIleIleAspGlyLeuSer 380
| | | | |
Db 1600 CTGTCCATGTCCGCGCCCGCAGCAGACGAGGTGTCAAGATTCATCATGCGCTCTCA 1659
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrr 400
| | | | |
Db 1660 GAGCAGAGAAACCGAAGAAAGAGATGCCAGCTGTGCTGATCCCGCCCATGCTGTAC 1719
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
| | | | |
Db 1720 GACGCTGACACAGCGGATCAAGTTCAATCAATGAACGGGCTTATGGCCCAACCCCATG 1779
QY 421 LysValTyrrLysAspArgGlnValMetAsnMetTPSerGluGlnGluLysGluThrPhe 440
| | | | |
Db 1780 AAGGTGTACAAAGACCGCCAGGTCTGAACATGTGAGAGACAGAGAGAGAGACTTTC 1839
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
| | | | |
Db 1840 CCGAGAAAGTTATGACAGATCCCAAGAACTTTGGCTGTATGTCATCTTCTGAGAGAG 1899
QY 461 LysThrValAlaGluCysValLeuTyrrTyrrLeuThrLysLysAsnGluAsnTyrrLys 480
| | | | |
Db 1900 AAGACAGTGGCTGAGTGGCTCTCTATTACTACAGTCAAGAAAGAAATGAAGAACTTAAG 1959
QY 481 SerLeuValArgArgSerTyrrArgArgArgGlyLysSerGlnGln---GlnGlnGlnGln 499
| | | | |
Db 1960 AGCTGTGTAGAGAGGAGCTATGCGCGCCGCGCAAGAGCAGAGNACAGNACGACGAG 2019
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
| | | | |
Db 2020 CAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAG 2079

[illegible]

Db	3108	GACGGGGAGGCGCGTGAAGCCACCGCGGAGGGGGCGCTCAAGGACAGAAAGAGGGGCG	3167
Qy	899	GLYSERGIYAAGLAATPHTHRAIALYSESERGIYALAPROGINAPSERAPSER	918
Db	3168	GGAGAGGCGAAGGCGCACCAAGCCCAAGAGCTCGGGCGCCCCCAAGACAGCGACTCACT	3227
Qy	919	ALAThrCYSESERIALAAGPGLUALAPGGLUALAGLUGLYGLYAPSLYSAHNRGLLEU	938
Db	3228	GCTACCTGCATGCAAGACGAGTGGATGAGCGCGAGGGCGCGCACAAAGAACCGCTGCTG	3287
Qy	939	SERPROARGPROSERLEULEUTHIR-PROTHGLYASPPROARGALAAANALASERPROGIN	958
Db	3288	TCCCCAAGGCGCACGCTCTCACCCCGACCTGGCGACCCCGGGCGCATGCTCACCCCGAG	3347
Qy	959	LYSPROLEUAPLEULYSGINLEULYSGINLAAGLALALALALALALALALALALALALAL	978
Db	3348	AAGCACACTGACCTGAAGACCTGAAGACACGAGCGAGCTGCATCCCTCCCATC--GTC	3404
Qy	979	THIRYSVALIHISGLUPROPROARGGLUALAPALALAPROTHIRLYSPROALAPROPRO-AL	998
Db	3405	ACCAAAAGTCCATGAGCCCCCGGGAGAGCAGAGCTCCACCAAGCAGACTCCCCAGAC	3464
Qy	998	APROPROBROPROGINASPLEUGINPROJUSERAPALAPROGININPROGLYSESE	1018
Db	3465	CCCCACCGCACCGCAAAACCTGACGCGGAGACGACGCGCCCTCAGACAGCTGGAGCAG	3524
Qy	1018	RPROARGLYLYSERIRGSEPROALAPROPROALAPSLYGLU-----1033	1033
Db	3525	CCCCGGGGCAAGAGCAGAGCCCGGACCCCCCGCGCAAGAGAGCAGAGACCTGT	3584
Qy	1034	-----ALAPHEALALAGLUALAGLUALAGLULYBLEUPROGLYASPPROCYTTRPH	1050
Db	3585	GTTCTTCCCAAGCTTGGCAGCCAGAGGCCAGAGCTGCTGGGGACCCCCCTTGCTGAC	3644
Qy	1050	ISEGGLYLEUPROBHEPROVALPROPROARGGLUALILLYLALASERPROHIALAPR	1070
Db	3645	TYTCGGGCGTGGCTTCCCGGTGGCCCCCGGTGAGTATCAAGGCTCCCGCATGGCCC	3704
Qy	1070	QASPPROSERALAPHESETRYRALAPROPROGLYHISPROLEUPROLEUGLYLEUHIAS	1090
Db	3705	GGACCCCTCAGCCCTTCTCCACAGCTCACCTGGTCAACCCAGTCGCCCTGGGCTTCATGA	3764
Qy	1090	PTHALAIRPROVALLEUPROARGPROPROTHIRISESERANPROPROLEUILESE	1110
Db	3765	CACGCGCCGGCGGCTCTGCGGCGCCACCCACCATCTCAACCGGCTCCCTCATCTTC	3824
Qy	1110	ISESERIALYHISAPROSERVALLEUGLUALRGGINILEGLYALILSESERINGLYMESE	1130
Db	3825	CTTGCCCAAGACCCCGAGGCTCTCGAAGAGGCAATAGGTGCATCTCCCAAGAAATGTC	3884
Qy	1130	RVALGINLEUHIHISVALPROTYISERGILHISALALYALAPROVALGLYPROVALTHME	1150
Db	3885	GGTCCACACTCAGAGTCCCGTAACAGAGATGCAAGGCGCGGCGGCGCTGCATCCAT	3944
Qy	1150	TGLYLEUPROLEUPROMETASPPROLYALYSLEULAPROPHESERGLYALIVSGHGL	1170
Db	3945	GGGGCTGCCCTGCGCATGAGCCCAAAAGGCTGGCACTTTCAGCGGAGTGAACAGAGA	4004
Qy	1170	UGINLEUSERPROARGLYGLNALAGLYPROPROJUSERLEUGLYVALPROTHALAGL	1190
Db	4005	GCAGCTGTCCCAAGGGGGCCAGGCTGGGCCACCGAGAGACTGGGGGGTGGCCACAGCCA	4064
Qy	1190	NGLUALASERVALLEULARGLYTHRALALEUGLYSERVALPROJGLYGLYSERLIEHLY	1210
Db	4065	GGAGGCGCTCGTCTAGAGGAGCAGACTTGGGCTCAAGTTCGGGGGGAAGATCACCA	4124
Qy	1210	SGLYILEPROSETHIRYVALPROSERASPPSERALALIEHTRYRARGGLYSERLIEH	1230
Db	4125	AGGCAATCCAGACACAGGGTGGCTTCGACAGCCCATCATACCCGGGCTCATATAC	4184
Qy	1230	THISGLYTHIRPROALASPPVALLEULYRGLYGLYTHIRLIEHTRARGLILEGLYGLUAS	1250

4185 CCACGCGACGCGACTGACCTCTGTACAAGGCGACCATCACAGGATCATCGGCGAGA 4244
QY pSerProSerArgLeuAspArgGlyValArgLysSerLeuProLysGlyValIleTyr 1270
Db CAGCCCGAAGTCCGTTTGACCGCGCGGAGAGACAGCTGCCAAGGGCCAGTCACTTA 4304
QY rGluGlyLysLysGlyValIleuSerTyrGluGlyMetSerValThrGlnCysSe 1290
Db CCAAGGCAAGAGGGCCAGTCTTGTCTATGAGGGTGGCATGTCTGTAGACCAAGTCTC 4364
QY rLysGlyLysArgLysArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgHis 1310
Db CAGGAGGAGCGCGAGAAAGAGCTCAGAGACCCCCCTGAGAGCGGCCGCCCAAGCGCAC 4424
QY rTyrAspMetMetGluGlyValArgValGlyValArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db CTATGACATGATGGAGGGCCGCGTGGGAGAGCACTCTCTCAGCCAGCATGCAAGTCT 4484
QY uMetGlyValArgAlaIleProProGluLysArgHisSerProHisHisLysLysGluGlnHis 1350
Db CATGGGCGGTGCATCTCCGCGGAGCGACAGCCCCCACCTCAAGAGCAGACCA 4544
QY rLysArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGlyLysArg 1370
Db CATCCGCGGAGTCCATCACCAAGAGATCCCTCGTCTCTACGTGAGAGCAAGAGGACTA 4604
QY rLeuArgArgGlyLysAlaLysLysLysLysArgGluGlyThrProProProProProSe 1390
Db CTGCGCTCCGGAGCGCAAGCTCTTAAGCGGAGGAGCAAGCTCCGCCCCACCGGCTC 4664
QY rArgAspLeuThrGlnAlaTyrLysThrGlnAlaGluGlyProLysLysLysProAl 1410
Db ACGGAGCATGACCGAGGCTTACAAGACGAGCGCTCGGCGCCCTGAAGCTGAAGCGGC 4724
QY HisGluGlyLysValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArg 1430
Db CATGAGGAGCTGTGTGGCCACCGGTGAAGAGCGGCGCTCTCATCATGATGATCCCGG 4784
QY gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluLys 1450
Db CGAGAGGCTGCGGCAACGCCCGAGCTGCCCTCGGCGCGCGCTCAAGAGGAGCTC 4844
QY rIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysI 1470
Db CATCAGCGAGGAGCCCGCTCAAGTACAGACCGGCGCTCCACCATGCTGCCCAAAA 4904
QY HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProle 1490
Db GACACGACGTACGCTCCCTCATCGCAGCGCCCGCGAGCGTCCACCGGTGACCGCT 4964
QY uAspValMetAlaAspAlaArgAlaLeuGluAlaArgAlaCysTyrGluGlyLysLeuLysSe 1510
Db GGAATGTGATGGCCGACGCGCGGCACTGGAACTGTCTGTAGAGAGAGCTTGAAGAG 5024
QY rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVal 1530
Db CCGGCGCAGGAGACCGCAGAGAGCTCGGAGGAGCTCATTTGCGCGCGCGCTCATTTG 5084
QY lProGluLysGlyLysProArgGlnSerProLeuThrTyrGluLysArgHisGlyValaProHis 1550
Db GCGTGTGCTGTGGAAACCGCGGACAGCCCTTAACCTATGAGAGACCAAGGCGACCTT 5144
QY eAlaGlyHisLysProArgGlyLysProValThrMetArgGluProThrProArgLeuG 1570
Db TGGCGGCGACCTCCACGAGGTTCCGCGGTGACACAGCGGAGGCCACCGCGGCTGCA 5204
QY nGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArg 1590
Db GAGGAGGAGGCTTTGCTGTCAGCAAGCATCCAGACCGAAGCTGACGTCGACGCTCG 5264
QY gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
Db TGAGATCGCCAAAGTCCCGCAGACACCGTGCCTGAGACCAACCCATTCCTCGCC 5324

1610 cTyrGluHisLysLeuLeuArgGlyValSerGlyValaLysLeuTyrArgSerHisIleProle 1630
Db CTATGAGCACCTGCTTCGGGCGGTGAGTGGCGTGAGACCTGTATCCAGCCACATCCCTCT 5384
QY uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyr 1650
Db GGCTTTGACCCCGACCTCATACCCCGGAGATCCCTCTGAGAGGACCGGCTGCTACTA 5444
QY rLeuProArgHisLysAlaLysProAsnProThrTyrProHisLysLeuTyrProProTyrLeuI 1670
Db CTTGCCCGGACACCTGTGGCCCCCAACCCACCTACCGGACCTGTACCCACCTACTAT 5504
QY eArgGlyTyrProAspThrAlaAlaLeuGluLysAsnArgGlnThrIleIleAsnAspTyrI 1690
Db CCGGCGCTACCCCGACACCGCGCGCTGGAGAACCGGACAGACCATCATCATGACTACTAC 5564
QY eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db CACTTCGACGAGATGACACACAGCGGCGCACCCGCTATGGCCAGCGAGTGTATGTCT 5624
QY uArgGlyLysSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG 1730
Db GAGGGGCTCTGCGCCCGGAGATCTCGTGGCATCTCACTACGCTGGGGTCCCGAGG 5684
QY yIleIleAspLeuSerGlnValProHisLysProValaLysValProProThrProGlyThr 1750
Db CATCATCGACCTGTCCAAAGTCCACACTGCTGTGCTGTGCCCCCGACACAGGCGAC 5744
QY rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
Db CCGACGACCGCGCATGAGACCGGCTTGTACTCTCCACCGGCGCCACCTTCACGAG 5804
QY rArgHisSerSerSerProLeuSerProGlyGlyProThrHisLysLeuThrLysProThr 1790
Db CCGCCACAGCACTCCCACTCTCCACAGAGTCCAAACACTGACAAACCAACCAAC 5864
QY rHisSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGlu 1810
Db CAGTCTCTGCTCGGAGCGGAGAGAGACCGGATGACAGCGGAGCGGAGATCGGAGG 5924
QY gGluLysSerIleLeuThrSerThrThrThrValGlnHisAlaProIleTyrArgProG 1830
Db GAAAAAGTCACTCTCATCTCCACACAGCGGTGAGAGCGACCCATGTGAGACTGG 5984
QY yThrGlnGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArg 1850
Db TACAGACGAGACAGCGGACAGCGAGCGAGCGAGCGCGGGGTGGGCGACGAGCGCG 6044
QY gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db CCCCCGCTCCCACTCCATGCCACACGACGACTCGCCCATCTCCCTCGGACCCAGGATCC 6104
QY lLeuGlnGlnArgProSerValaLysHisAsnThrGlyMetLysGlyIleIleThrAlaVal 1890
Db CTTCCAGCAAGAACCCAGTGTCTTCAACAACAGGCAAGAGGATATATACACGCTGT 6164
QY lGluProSerLysProThrValaLys-----ArgSerThrSerThrSerSerProVal 1907
Db GAGAGCCACAGCCCGACAGGCTCTGAGGAGGAGGAGGCTCACTCCACTCCGACCCGT 6224
QY lArgProAlaAlaThrPheProProAlaThrHisCysPProLeuGlyGlyThrLeuAspG 1927
Db TCGCCCGGCTGACACTTCCCACTTCCCACTTCCCACTTCCCACTTGGGGGAGCACTTCATGG 6284
QY yValTyrProThrLeuMetGluProValaLysLeuLeuProLysGluAlaProArgValaAla 1947
Db GGTCTACCTTACCTTATGAGCCCGTCTTGTGCTGCGCAAGAGAGGCCCGGGGTCCCG 6344
QY gProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
Db GCCAGAGCGGCCCCGAGAGACAGCCGGCCATGCTTCTCTGCGCAAGCCCCGAGCGGCTC 6404

QY 1967 rglyleugluproalaaserseproserlysglysergluproargproleuvalpropr 1987
 Db 6405 CGGGCTGAGACCGCGCTCTCTCCCAAGAGGCTCGAGGCCGCGGCTTAGTGGCTCC 6464
 QY 1987 ovalserglyhiabalahrtlleaiaargthrproalalyabamleuvalprohishsal 2007
 Db 6465 TGTCTGGGCGACGCCACATCGCTCGACCTCGCAAGAACCTCCACCTCACCGC 6524
 QY 2007 aserproaspbroproalaproproalaserasaseraspbrohisarggluyserthrgl 2027
 Db 6525 CAGCCCGGACCGCGCGCGCACCTGCTCGGCTCGGACCGGACCGGAAAAAGCTCA 6584
 QY 2027 nserlyseprophaserleegngluleugluleuargserleuglytrhiisgliserse 2047
 Db 6585 AAGTAAACCTTTTCCATCCAGGAACCTGAACTCCGTTCTTGAGGTTACCAAGCGCACAG 6644
 QY 2047 rtyrserprogluglyvalgluprovalaserprovalaserproserleuthrhias 2067
 Db 6645 CTACAGCCCGAAGGGGTGAGCCCGCTCAGCCCTGTAAGCTCACCCAGTCTGACCCACGA 6704
 QY 2067 ptyrsglyleuprolyshisleugluleuaplyserhisleugluleugluleuarg 2087
 Db 6705 CAAGGGGCTCCCAAGACCTCGAAGAGCTCGACAAAGCCACTTGAGGGAGGCTCGG 6764
 QY 2087 gprolysglnproglyprovalylyleuglygllyalalahiisleubprohisleuar 2107
 Db 6765 GCCCAAGACAGCAGGCCCGGTGAAGCTTGCGGGGAGCGCCGACCTCCACACCTCGG 6824
 QY 2107 gproleuprogluserglinproserseproleuileuglithralaprogllyalily 2127
 Db 6825 GCGCGTCCCTGAGAGCCAGCCCTGTCAGCCCGCTGTCAGACCCCGCAGGGGTCAA 6884
 QY 2127 sgllyhisglinarvalvalthrleuaglnhiisleserglvaliljethrglnappty 2147
 Db 6885 AGGTACACAGCGGGGTGTCTACCTCGGCCACAGCATAGAGGTCTACACAGGACTA 6944
 QY 2147 rthraighishaproglnleuuseralaproleuproalaproleutyrserepnr 2167
 Db 6945 CACCCGGACACCCACAGCAGCGCTGACGCCCTCGCCCGCCCTTACTCTCTCC 7004
 QY 2167 oglyalaserctyprovalleuapleuargtargproproseraspleutyrieupr 2187
 Db 7005 TGGGGCAGCTGCCCCGTCTCTGAGACCTCCCGCCACCAAGTGAACCTTCCCGCC 7064
 QY 2187 opromaphisglialaproalarglyserprohisasersegllygllylsarserpr 2207
 Db 7065 CCGGACCATGTGTCCCGCGCGCTGCTCCCGCACAGGAGGGGACAGGTCTCC 7124
 QY 2207 oguproasnlysthrsevalleuglygllygllyuaspglyleugluprovalserpr 2227
 Db 7125 AGAGCCAAACAAACGTCGCTTGGGTGTGTGGAGGACGGRATTGAACCTGTCTCC 7184
 QY 2227 oproglyglymetthrghluproglyhiisaserargseralavaltyrproleuleutyar 2247
 Db 7185 ACCGGAGGGGATACCGAGCCAGGACCTCCCGAGTGTGTACCGCTGCTGTACCG 7244
 QY 2247 ghaepglyglugluthrghluproserargmetglyserlyserprogllyanthsiergi 2267
 Db 7245 GATGTGGGAACACAGCGAGCCACAGAGATGGGCTCCAGTCTCCAGGCAACCGACCA 7304
 QY 2267 nproproalapheserlyshleuthrgluserasaseralamevallysserlyshly 2287
 Db 7305 GCGGCGACGCTTCTTACGAAAGCTGACCGAGCAACTCGGCATGGTCAAGTCAAGAA 7364
 QY 2287 sglnglulileuamlyleuamthrhisaamargamgluproglyutyranthiese 2307
 Db 7365 GCAAGAGATCAACAAAGAGTGAACCCCAACCGGAATGAGACTGAAATACATATCAG 7424
 QY 2307 rglinproglythrglulileuammetproalailjethrglythrglyleuamethrty 2327
 Db 7425 CCGAGCTGGGACGGAATCTTCAATATGCCCGCATACACGGAAACAGGCTTATGACCTA 7484
 QY 2327 rargserglinalvalglnglunhiisaserthrasmetglyleuglunlaileilear 2347

Db 7485 TAGAGCGACCGGTGACGAGAACATGCACACCAACATGGGGCTGAGGCCATTAATTAG 7544
 QY 2347 glyvalaleuamethglylyethrtyrargntprglugluserproproleuserlisaenal 2367
 Db 7545 AAAGGACCTATGATGTAATATATGACAGTGGAGAAATCTCCCGCTCAGCGCAATGTC 7604
 QY 2367 apheasnproleuamshlaseralaserleuproalalameproliethralalalae 2387
 Db 7605 TTTTAACCTCTGAATCCAGTCCAGCTCGCCGCTGCTATGCTCAATACCGTGTGA 7664
 QY 2387 pgllyarserasphishrleuthrserprogllygllygllylyvalyvalysergi 2407
 Db 7665 CGGACGAGAGAACACACACTCACTCGCAGGTGGCGCGGAGGAGCCAAAGTCTCTGG 7724
 QY 2407 yargprosersearglyvalalyserproalaprogllyleuamshlaserllyasparpr 2427
 Db 7725 CAGACCGACGCGGAAGAACCAAGTCCCGCGCGCGGCTGATCTGGAGACCGGCC 7784
 QY 2427 opromerlyasersevalhissergluglyasparcyasamargargthrproleuthras 2447
 Db 7785 ACCCTGTCTCTTCACTGCTGCTCGAGGAGACTGCACCGCGGACGCTCACCA 7844
 QY 2447 natgvalttrpgluaspargprosersearglyserthrprophetrotyranprole 2467
 Db 7845 CCGCGTGGGAGAGACAGGCGCTGTCGCGAGGTTCCAGCCATTCCTTACAAACCCCT 7904
 QY 2467 ullemethargleuglnalaglyvalmethalaserproproproprogllyleuproalagl 2487
 Db 7905 GATCATGCGGCTCAGCGGGGTGTCAATGCTTCCACCCCAACCGGCTCCCGCGG 7964
 QY 2487 ysergllyproleuamgllyprohisishlaltapabglugluprolyasproleuency 2507
 Db 7965 CAGCGGCGGCTGCTGCGGCGCCCAACAGCCTGGAGAGAGAGACCAAGCACTGCTCTG 8024
 QY 2507 sserglnyrgluthrleuserasparserglu 2517
 Db 8025 CTCGAGTACGAGACACTCCGACAGCGAG 8055
 RESULT 4
 AF113001 8388 bp mRNA linear ROD 20-MAR-1999
 LOCUS
 DEFINITION Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
 ACCESSION AF113001
 VERSION AF113001.1 GI:4454547
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 REFERENCE 1 (bases 1 to 8388)
 AUTHORS Ordentlich,P., Downes,M., Xie,W., Genin,A., Spigner,N.B. and Evans,R.M.
 TITLE Unique forms of human and mouse nuclear receptor corepressor SMRT
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
 MEDLINE 99178941
 PUBMED 10075563
 REFERENCES 2 (bases 1 to 8388)
 DOWNES,M.R., ORDENTLICH,P. and EVANS,R.M.
 TITLE Direct Substitution
 JOURNAL Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source location/Qualifiers
 1..8388
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="spleen; brain"
 626..8047
 /function="transcriptional co-repressor"
 /note="mSMART alpha"

```
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid
hormone receptor alpha"
/protein_id="AAD20944.1"
/db_xref="GI:4454548"
/translation="MSGTQPVAAQMTRAAPRYPHGISYRVPVQIARSHTDVGLLEYOH
HPRDITSHLSFGSSTIQPQRSPRLSISEPQSGSEBTLKDRSLAGLEPVSPSPHAPDEL
TSKRPRLPLDPTLRPSPLATQPSGSEBTLKDRSLAGLEPVSPSPHAPDEL
ELAPRLSKBELIQMDRVREITWVEQOISLKKKQOQLEBEAARPEPEKVPSP
IESKRSILQIYDENRKAFAAHRIEELGQVLELPLQPSDITQVYENKIKQAM
RKLLIYFKRRHARQWQRCORQDOLMAEMKEVEEIKENNPRAEVEKREYRE
KQPELRKQRELOEMQSRVQSGSLSMARSEHVEIIDGISEQENLEKQREY
AVTPMLYDADQRIKFTMMGLMDPMKYVYKDRQVTTMWSQERDPTPEKQWQPKA
FGLIASFERKTVAECLVYLYLTKKNENTKSLVRSYRRGRSQQQOQOQOQOQOQ
ASSQBEKEKEKEKADBEKODABEKSLSEKTDGTGSDHEKEAASGKRG
TANSQGRKRIITRSMANENHETATPOOSSELSAMENNESRPTMEEMETAKGL
BHGMSALIAWVQSKTVGCOQNFYNYKRONLDEILQOHLKMEKEMARKKKKT
PAASEETAAPPAARDEMEWASGASNEELAEAAQSASNEUPVVECGSPAAV
NSDTESVSPSPSEAMKDTGKPTGTETALPAATQPPVPEEPAAPAPSPVPDASG
PSPSPSHLPHRLMTMKNKPKRLQIQRMPKSLRPRSRMEPEEPEASEE
PSPSVSDHEKETEEDKAGTEALIVTSEAPLKEBAGSKAAVTKSSGATQDS
DESATCSADEVDEPEGDKRLSPSPSLTPAGDPASTSPQKPLDLKQLKQRAAI
PIQVTKSEPREDTVPKVPVPPROHLOPEGDVQOQSGSPRGRSPVPAP
KEAKRPAAPAPPTGPKLPTPPMWSGLPPPIPREVITKTSPPHADSASPTPG
HELPLGHDASAPVLPRIISNPPLISAKHPGVLRLGAIQQQMSVQLRVHSE
NAKMPGLTWGLPLAVDPKUGTALGATSGITKGLPSTPADSPSTRSITHTGP
ADVLKGTISRTVGEDSPRLDRAREDTLPKHVLYEGKGVHLSLEGMSVQSCKE
DORSSSGPHETAAKRTYDMMGKRGVTVTSASIEGLMRAIPEQSHPLKQHR
GSITQIPRSYEAQEDYLREAKLKEKGTGPPPPDLTETVYKRLPLDGPKL
KPTHGCVATSEKAGRSIHEIPREBELRPBLIAPRLKESITGQPTLKYDQSGPS
TGTKGDVASTIGSPRRPALHPDIDMAALBACREBSKSKSSGTSAGSAGSIT
RQAPVVPPELGRKOSPITYEDHAPFTHLPGSPVTTRETPRLQESLSSKASQ
DKLTSTPREIAKSPHSTVPEHNPRISEYHLKRVTVDLVYRGIPLAPFTSLPR
GJPLEAAAYLPHLAPSPYPLPYLIRGYDPTALENRQTIINDYITSQOM
HNAASMAQADMLRGLSPRESSIALNVAAPRGIDLSQVHLPLYVPTGTPATA
IDRLAYLTPAPPSRSHSSPLSGCPHILAKPTATSSERREBERERDSTILST
TVEHAPITWRPCTGSSGAGSSRPASHTHQSPISPRQDALQAPSTIANHSMKV
VTSVEPTVLRAWASTSTSSVPRAAPFPATHCPLGCTLEGVPTIMEVLPKE
TSRVARERARVADAGHAFLLKPRGPAPSPSKSEBRPILAPSPSHTILATPANK
AHHNASDPAPSTASDLREKTQSKPFSIOELIELSLRSLAGSPGVPISVPS
SPLTNDKLSLEBELKSHLGEGLRHQOPGMKLSAANLPHLRPPEOPSSP
LIOTAPGIGHORVYLAOHISVITQDPTBHHPOOLSGYLPSPGACCPVLD
RRPBDLYLPDPHGTTPAGSPHSEGGKSPSPSTSVLGSSEDAIEPSPREGNTER
GHARSTAYPLLRDGEQGRMGLSPGNTSQPTEFSKLTSSNMAVSKQEOENK
LWTHRNBEYNIQGPTEI FMWPAITGALTCRQAAVDSHALSTPGCGAKYSGR
GKDWDEBPPLGANAFPLNASALTPAAMPITTAGDSOHAISTPGCGAKYSGR
PSSRKSPAPGASGDRPVSYSVSEDCNRRTPLTRWVEDRVSAGSPVPYNP
LMRLQAGYMASPPPGGLAAGGFLAHPHMADEBPKLLCCQYETLSSE"
```

711..1187
/note="Region: deleted in SMRT beta isoform encoded by
GenBank Accession Number AF113002"

misc_feature

ORIGIN

Alignment Scores:

Pred. No.:	2, 74e-215	Length:	8388
Score:	11067.00	Matches:	2160
Percent Similarity:	88.20%	Conservative:	89
Best Local Similarity:	84.71%	Mismatches:	192
Query Match:	83.75%	Indels:	110
DB:	10	Gaps:	29

US-09-522-753-5 (1-2517) x AF113001 (1-8388)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20

Db 626 ATGTCAAGATCCACACAGCCTGTGGCACAACATGCGGGCTCTGACCCCGCTACCA 665

QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40

Db 686 CCCCATGATCTCTTCAACCGGCTGCAATAGCCCGCTCCACACGAGCGGGCTGCTT 745

QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60

Db 746 GAGTACCAACACACCCCGGTGACTACACTACACCTGTACACCCGGTTCATCATCCAG 805

QY 61 ProGlnArgArgProSerLeuLeuSerGlnProGlyAsnGluArgSerGln 80

Db 806 CCACAGAGGAGCGGCCCTCACTGCTGTACAGTTCCAGCTCGGAGAAAGGCTTCCAG 865

QY 81 GluLeuHisIleLeuArgProGluSerHisSerTyrIleuProGluGluGlyArgIleuMet 100

Db 866 GAGCTCCACCTGCGCCCTGAGTCCCGACAGTTCCTGCTGAGCTGGGCAAGCCCGACATA 925

QY 101 GluPheIleGluSerIleArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120

Db 926 GAATTACCGAGACGACAGAGCCCGCCCTGAGACTACTACCGGATACCTGTGCGGCCA 985

QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140

Db 986 TCACCCCTGTGCGCACCTGGACGAGGTGGTCTGAAGACTTAACCAAGACCGCTAGC 1045

QY 141 LeuThrGlyIleLeuGluProValSerProProSerProHisThrAspProGluLeu 160

Db 1046 CTGGCAGGACAGCTGAGCTGTGTCACTCCAGTCCCGCAGCTGACCTGAGCTA 1105

QY 161 GluLeuValProProArgLeuSerIleGluGluLeuIleGlnAsnMetAspArgValAsp 180

Db 1106 GAGCTGGCGCCCATCTGCACTGTCCAAGAGAGAGCTGATCAGAAATGAGCCGCTGAGC 1165

QY 181 ArgGluIleThrMetValGluGlnGlnIleSerIleLeuLeuIleGlnIleGlnIleu 200

Db 1166 CCGTAGATACCATGTGTGAGAGCAGAGATCTCAAGCTGMAAAGACAGCAACGCTTG 1225

QY 201 GluGluGluAlaIleAlaIleProProGluProGluIleProValSerProProIleGlu 220

Db 1226 GAGGAGAGAGCGCCGACAGCGCCGACGAGAGCTGTGTCCGACACCCCATAGAA 1285

QY 221 SerIleHisIleArgSerLeuValGlnIleIleTyrAspGluAsnArgIleIleGluAla 240

Db 1286 TCMAAGCACCGAAGCTGCTCAGATCATCTGATGAGAAACGGAGAAAGACCGAAGCC 1345

QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGlnPro 260

Db 1346 GCACACCGGATCTTGAAGAGCTGGGGCCCGAGGTGAGACTGCTCTGTACCAACGACCG 1405

QY 261 SerArgThrArgGlnTyrHisGluAsnIleIleValIleAsnGlnAlaMetArgIleValLeu 280

Db 1406 TCTGACACACGCGCAGTACATATAAAATCAAAATTAACACAGCCGATGGAGAAAGCTG 1465

QY 281 IleLeuTyrPheIleArgArgAsnHisIleAlaArgIleGlnTyrIleGlnIlePheCysGln 300

Db 1466 ATCTTGTTACTTTTAAGCGAGAGAACACCGCGCCGACAGCATGGGAAACAGGCTTCTCCAG 1525

QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluIleValIleValIleGluAsnAsnPro 320

Db 1526 CGCTATGACCAAGCTCATGAGGCGGTGGAGAAAGGTAGAGCGCTATGAGAAACAATCCG 1585

QY 321 ArgArgArgAlaIleGluSerIleValArgGluTyrTyrGluIleGluGlnPheProGluIle 340

Db 1586 CGAAGAGGCGCAAGAGAGCAAGGTGAGAGGTACTAGAAACAGTTCCTCGGAGATC 1645

QY 341 ArgIleGluArgGluLeuGlnGluArgMetGlnSerArgValGlyIleArgIleSerGly 360

Db 1646 CCGACGACGCGGAGCTGACAGAGCCGATGACAGAGAGGCTGGCCAGGCTGCGACTGGG 1705

QY 361 LeuSerMetSerAlaIleArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380

Db 1706 CTCTCCATGTCGGCTGCCCGCAGTGAAGTATGAGGTTTCTGAGATCATTAATGCTGTGCT 1765

QY 381 GluGlnGluAsnLeuGluIleGluGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400

Db 1766 GAGCAGGAGAACCTGAGAGAGAGATGAGCGCCAGCTGGCCGTATCCGCCCATGTTGAC 1825

QY 401 AspAlaAspGlnGlnArgIleIleIlePheIleAsnMetAsnGlyLeuMetAlaAspPromet 420

Db 1826 GACGCGGACCAAGAGATCAAGTTCAATCAATGAATGACTCATGATGATGACCCCATAG 1885

QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnLysGluThrPhe 440
DB 1886 AAGGCTTACAAGACCGGTCAAGGTTACCAACACTGTGGACGAGCAGAGAGACACTTC 1945
QY 441 ArgGluLysPheMetGlnHisProLysAsnDheGlyLeuIleAspPheLeuGluArg 460
DB 1946 CGTGAGAGATTATGACAGACCCCTAAGAACTTGTGCTGATGCTCATCTCTGAGAGA 2005
QY 461 LysThrValIleGluCysValLeuTyrTyrTrpLeuThrLysLysAsnGluAsnTyrLys 480
DB 2006 AAGCGGTGCTGTGATGTGTCTTATTTACTGACCTGACCAAGAAAGAAATTTACAG 2065
QY 481 SerLeuValArgAspSerTyrArgArgArgGlyLysSerGlnGlnGlnGln 500
DB 2066 AGCTGTGTAGGCGGAGACTATCGGCGCGGTGGCAAGCCGACGACGACGACGACA 2125
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
DB 2126 CAACAGCAGCAGCAGCAGCAG-----ATGGCAGCAGCAGCAGCAGAGAGAG 2173
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysProGluValGluAsn 540
DB 2174 GAGGAG 2233
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAspGlu 560
DB 2234 GAG 2293
QY 561 LysGluLysValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGly 580
DB 2294 AAGAGAGCGCTGGCTCCAAAGCGCCGAAACTCCCAAGCAGCGCGCGCGAAAGGC 2353
QY 581 ArgLLeuThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 600
DB 2354 CGATATCGGCTCTCATGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2413
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlnGluMet 620
DB 2414 AGTTTCAGAGCTGCTTCATGAGATGAAACAGAGTTCTCGCTGACTGAGAGAGAGATG 2473
QY 621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
DB 2474 GAGACAGCAAGAAAGGCTCTGTGAACATGGAGAGAACTGTGACCATTTGCCGATG 2533
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysValArgGln 660
DB 2534 GTGGGCTCCAAAGACCGGTGCCAGGTGAAGAACTTCACTTCACTCAAGAAAGAGCAG 2593
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluValArgAsnAlaArg 680
DB 2594 AACCTGACCAAAATCTTTCAGACGACACAGCTAAAGATGAGAAAGAGAGAGAGAGCTGG 2653
QY 681 ArgLysLysLysLysValProAlaAlaAlaSerGluGluAlaIlePheProProVal 700
DB 2654 AGGAAAGAAAGAAAG 2713
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
DB 2714 GAG 2773
QY 721 AlaGluLysLeuHisLysSerGlyAsnGluValProArg---GlyGluCysSerGlyPro 739
DB 2774 GCGAAGAGCTCACAGGCTCTGGGAATGAGAGTTCCCAAGTTGGAGAGAGAGAGAGAG 2833
QY 740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla 759
DB 2834 GCTGCTGTCAACAACAGCTCTGATCTGAGAGTGTCCCATCCCGGCTTCAGAAAGCATG 2893
QY 760 LysAspThrLysGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAspGly 777
DB 2894 AAGGACACT-----GGGCTTAAACCACTGACGTGACGTGAAGCATTTGGCGTGCCACC 2944

QY 778 ProProGluGlyProProThrProProArgArgThrSerArgAlaProIleGluProThr 797
DB 2945 CAGCCACTGTCTCT-----CTCCAGAAAGAACCGGACAGTATGCGCTGAGACCTCC 2998
QY 798 ProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAla 817
DB 2999 CCAATCCCTGATGACGTGGCCACATCCCCAGAGCTTC---CCATCATCTGCGCGCA 3054
QY 818 ProProProValValProLysGluGluGluGluGluGluThrAlaAlaProProVal 837
DB 3055 CCCCCGCTACTGTGACAAAGATGAACAAAGAACCCCCGGCTGCTCCAGCTCCCAAGACA 3114
QY 838 GlnGluGluGluGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLys 857
DB 3115 GAGGATGCCAAGAGAGAGAGAGTGTGAGCCGAGAGAG-----ATCGATGTGGAAAA 3165
QY 858 Ala-GluGluPro-----ValLysSerGlyCysThrGly 868
DB 3166 GCCAG 3225
QY 868 uGluLysGluGluGlyProAla---LysGlyLysAspAlaGluAlaIleGluAlaThrAla 887
DB 3226 GAGAACCGAAGAGAGAGAGCTTAAGACAAAGCCAGAGGACAGAGGCTATTGAACGTGTGTC 3285
QY 887 AGUGLYAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThr---Al 906
DB 3286 TGAAGCACCACTTAAGGTGAG-----GAGGCTGTATGCAAGGACACTGTGACCAAGG 3339
QY 906 aLysSerSerGlyAlaProGluAspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
DB 3340 TTCCAGCTCAGGTGCCACCCAGAGCAGTGACTTCAGTGCCACCTGACGTGCCGATGAGT 3399
QY 926 LaAspGluAlaGluGlyGlyLysAspLysAsnArgLysLeuSerProArgProSerLeuLeuThr 946
DB 3400 GAGAGAACCCGAAGAGAGAGTGAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3459
QY 946 rProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLe 966
DB 3460 CCGGCTGAGATCTCCGGGCGAGTACTGCTGCCAGAGAACCGCTGAGACTGAAGAGAGCT 3519
QY 966 uLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArg 986
DB 3520 GAAGCAGCGAGAGAGCGGCTATCCCTATCCAGGTCAACAGAGTCCATAGAGCCCGCCG 3579
QY 986 GGIAspAlaAlaIleProThrLysProAlaProProAlaProProProGluAsnLeuGlu 1006
DB 3580 GAGAGACACAGTACCCCAAGAGCCAGTCTCCCTGTGCTTCCACCCAGCAGACCTTACA 3639
QY 1006 nProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPr 1026
DB 3640 GCCAGAGAGTGAAGTGTCTGACAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3699
QY 1026 oAlaProProAlaAspLysGlu-----AlaPheAlaAlaGlu 1038
DB 3700 AGTGCTCTCTGCGAGAAAGAGCAGAGAAACCCGACTTCTTCGCGCTTCCCAACTGA 3759
QY 1038 uAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValPr 1058
DB 3760 GGGCCCAAACTTACCACTGAGGCCCAAGCTGTGATCGGCTGCTGCTTCCCACTCC 3819
QY 1058 oProArgGluValIleLysValaSerProHisAlaProAspProSerAlaPheSerTyrAl 1078
DB 3820 TCCACGGAGAGTATCAAGACTTCCCAACAGCGCGCTGACCTTCGCTTCTTCAAC 3879
QY 1078 aProProGlyHisAspProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProAr 1098
DB 3880 ACCCGCGGTCAACCCCTGCTGTGGGCTTCCAGTATGAGCCCGGCTCTCTGCGCAGC 3939
QY 1098 gProProThrIleSerAsnProProProProLeuIleSerSerAlaLysHisProSerValLe 1118
DB 3940 TCCCGCC---ATTCTTAACCCCAACCCCTCATCTTCTTGGCCAAACATCTCCGGGCACT 3996
QY 1118 uGluArgGlnIleGlyAlaIleSer---GlnGlyMetSerValGlnLeuHisValProTy 1137


```

Db      3997  TGAAGAGGAGCTGGGTGCTATCCACAGCAGGGGAGTGCATGCCAGCTTCGATGGCTCA 4056
Qy      1137  rSerGluH1sa1a1yVala1ProVala1GlyProVala1ThrMetGlyLeuProLeuProMetAs 1157
Db      4057  CTGAGCAGTGGCCCAAGGCCCTTCCTTACCATGAGGGCTCCCTGGCTGGA 4116
Qy      1157  pProLyS1yLeuA1a1ProPhaSerGlyVala1ySGInGluInLeuSerProArgGlyG1 1177
Db      4117  CCTTAAGAAGCTG----- 4129
Qy      1177  na1aG1yProProG1uSerLeuGlyVala1ProThra1aGInG1ua1aSerValLeuArgG1 1197
Db      4130  -----GG 4131
Qy      1197  yThra1a1eug1ySerVala1ProG1yG1ySer11eThrySG1y11eProSerThraArgVa 1217
Db      4132  GACAGCACTGGGCTCCGCCACACAGTGAAGCATCACCAAGGCGCTCCCACTACCCAGGCG 4191
Qy      1217  1ProSerAerSerA1a1eThrTyzaRg1ySer11eThrH1aG1yThrProH1aAspVa 1237
Db      4192  TGCAGAGGGCCCC-----AGCTACAGAGGCTCTATACCCACGCGACCCCGCAGACCT 4245
Qy      1237  1LeuTyrySG1yThr11eThraRg11e11eG1y1uaerSerProSerArgLeuAspAr 1257
Db      4246  CCTCTCAAGAGGTACATCAGCAGATCGTCGGTGAGAGCAGCCCAAGTCGCTTGACCG 4305
Qy      1257  gG1yArGg1uaerSerLeuProLySG1yH1aSVa11eTyrg1uG1yLyVa1ySG1yH1aSVa 1277
Db      4306  GGCACAGAGAGACCCCTGCCCAAGGGCCATGTCATCTAGAGGGCAAGAAAGGCCACGT 4365
Qy      1277  1LeuSerTyrg1uG1yG1yMetSerVala1Thrg1nCySerLySG1yAspG1yArGserSe 1297
Db      4366  CCTATCTTAAGAGGTATGTCCGTGCACAGTCTCTTAAGAGAGATGGAAGAGCGAG 4425
Qy      1297  rSerG1yProProH1aG1uThra1a1aProLySaRg1ThyArpMetMetG1uG1yAr 1317
Db      4426  CTCGGGCCCAACCCATGAGACTGCCCCCTTAAGCCACTATGACATGATGAGGGGCG 4485
Qy      1317  gVala1yArG1a1a1eSerSerA1a1eG1uG1yLeuMetG1yArG1a1a1eProPr 1337
Db      4486  TGTAGGACAGAGCTGTACCTCAGCCAGCATAGAGGACTCATATGGCCGGCCATC---CC 4542
Qy      1337  oG1uaerH1aSerProH1aH1eLeuLySG1uG1nH1aH1eArG1ySer11eThrg1 1357
Db      4543  TGAAGCAGCAGGCCCC---CACTCAAGAGCAGCATCATCCGAGGCTCCATCAGCA 4599
Qy      1357  nG1y11eProArGserTyryVala1G1ua1aG1nG1uaerTyryLeuArG1ua1a1yS1e 1377
Db      4600  AGGCAATCCGAGGGTCTATGTGAGGGCGCAGAGACTACTTACGGCGGAGGCCAAGCT 4659
Qy      1377  uLeuLySaRg1uG1yThrProProProProProProSerArgArG1eThrg1ua1a1y 1397
Db      4660  CTTGAAGCGAAGAGGACACACCCACCCACCACTCGGAGCTGATGAGACCTA 4719
Qy      1397  r1ySerThrg1n-----Ala1eug1yProLeuLyS1eLySProA1aH1eSG1uG1y1e 1414
Db      4720  CAAGCCCGGCCCCCTGGACCTCTGGGTCCCTGGAAGCTGAAGCCAGACCAAGGGGTG 4779
Qy      1414  uVala1a1Thra1yVala1yG1ua1aG1yArGser11eH1eG1u11eProArG1uG1uLeuAr 1434
Db      4780  GGTAGACAACTGTGAAGAGAGCGGCGCTTATCTATCATGAGATCCCAAGAGAGAGCTGGG 4839
Qy      1434  gH1aThrProG1uLeuProLeuA1a1ProArGProLySG1uG1ySer11eThrg1nG1 1454
Db      4840  CCGCACAACCTGAGCTACCTCGCACCAAGGCTCTGAAGAGAGGTTCATCAACCAAGGG 4899
Qy      1454  yThrProLeuLyTyryAspThrg1yAlaSerThrThrg1ySerLyS1eLyS1aAspVa1aR 1474
Db      4900  CACCCCACTTAATAGACTCTGGGGCACTCCATCGGACCAAGAAACACACACCTGGG 4959
Qy      1474  gSerLeu11eG1ySerProG1yArGThrPheProProVala1H1aProLeuAspVa1aMetAl 1494

```

```

Db      4960  CTCATCATTCGGCAGCCCCCGGCGCTTTCCTGCTCCCTGCACCCGCTGGACATTAATGGC 5019
Qy      1494  aAspA1aArG1a1a1eug1uaRg1a1CyEtyrg1uG1uSerLeuLyS1eArGProG1yTh 1514
Db      5020  TTAACCCCGGGGCACTGGACCGTCTGTCTTAAGAGAGTCTGAAGAGCGGGTCAAGGAGC 5079
Qy      1514  rAlaSerSerSerG1yG1ySer11eAlaRg1yAlaProVala11eVala1ProG1uLeuG1 1534
Db      5080  CAGCAGTGTGTGAGGGGCTTCATCACAGTGGGCTTCAGTGTGTGTCTGAACTGGG 5139
Qy      1534  yLySProArGg1nSerProLeuThrTyrg1uAspH1aG1yAla1aProPha1aG1yH1aSe 1554
Db      5140  CAAGCCAGGCAAGCCCACTGACTTACGAGACCAAGGGGACCTTACCAAGTACCT 5199
Qy      1554  uProArG1ySerProVala1ThrMetArG1uProThrProArGLeuGInG1ySer1e 1574
Db      5200  GCCAGTGGCTCCCTGTGTACACAGAGGAGCCAGCCAGGCTTCAAGAAAGGACCT 5259
Qy      1574  uSerSerLySa1aSerG1nAspArG1yLeuThrSerThrProArG1u11eAla1y 1594
Db      5260  CCTATCCAGCAAGGGGTCCAGAGACCGGAACTGACATCTACACCCCGGAGATCGCCA 5319
Qy      1594  sSerProH1aSerThra1ProG1uH1aH1aProH1aPro11eSerProTyrg1uH1aSe 1614
Db      5320  GTCCCAACACAGCACTGTGCCAGACACACCTCACCCATCTCCCTTATGAGCACTT 5379
Qy      1614  uLeuArG1yVala1SerG1yVala1a1aPLeuTyryaRgserH1a1eProLeuA1aPheAspPr 1634
Db      5380  GCTCCGGGGCGTGAAGTGTGTGACCTGTACGTTGTGCATCCATTCGCTTGACCC 5439
Qy      1634  oThrSer11eProArG1y11eProLeuAsp---Ala1a1a1a1a1yTyryLeuProAr 1653
Db      5440  CACTTCATACCCCAAGGAGTCTCTGGAACAGAGCGGAGCGCTTACTTCCGCG 5499
Qy      1653  gH1eLeuA1aProArProThrTyryProH1aLeuTyryProProTyryLeu11eArG1yTy 1673
Db      5500  GCACTTGAGCCCAAGCCCACTACCCACACCTGTACCACTTACTTATCCGCGCTA 5559
Qy      1673  rProAspThra1a1a1eug1uaRg1nThr11e1eAspTyry11eThrSerG1 1693
Db      5560  CCTGACACGGGGCGCTTGAGACCGCCAGCAACATCATTAATGACTATCATCCGCA 5619
Qy      1693  nG1nMetH1aH1aThra1a1a1eThra1a1eThra1a1a1eThra1a1a1eThra1a1e 1713
Db      5620  GCGAGTGCACCAACAGCTGCTCCGCGCAATGCGCAGCTGTGACATCTGAGGGGT 5679
Qy      1713  uSerProArGg1uSerSerLeuA1a1eAspH1a1a1aG1yProArG1y11e11eAs 1733
Db      5680  GTACCCGAGAGTCTCTGCTGCTCAATTAATGCGCTGGCCCAAGAGCATTAATCA 5739
Qy      1733  pLeuSerG1nVala1ProH1aLeuProVala1eUVala1ProProThrProG1yThrProA1aTh 1753
Db      5740  CTTGTCCCAAGTGCACACCTGCTGCTGTGACCAACAGCCAGGACCCCTGCCAC 5799
Qy      1753  rAlaMetAerArG1eUa1a1yTyryLeuProThra1a1aProG1nProPhaSerSerArgH1aSe 1773
Db      5800  CCGCAATCGACCCGCTTGTCTTACTCCCTCCACTCGCCGCCACCTTCACAGACGCGCACAG 5859
Qy      1773  rSerSerProLeuSerProG1yG1yProThrH1a1eThrThryLeuThryProThrThrThrSe 1793
Db      1793  rSerG1uaRg1uaRgAspArG1uaRgAspArG1uaRgAspArG1uaRg1uLyS1e 1813
Qy      5860  TTAGCTCAACGCTGTCCCAAGAGGCCCACTCACTAGCTTAACCAATGCCACATCTTC 5919
Db      1793  rSerG1uaRg1uaRgAspArG1uaRgAspArG1uaRgAspArG1uaRg1uLyS1e 1813
Qy      5920  ATCCGAGCGGGAACGGGAACGTGAGCGGAAGAGC-----AAGTC 5961
Db      1813  r11eLeuThrSerThrThrThra1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 1833
Qy      5962  CATCTCAAGTGTACACATGAGTGAAGATCACCACTGTGAGACCTGTGATACGAGCA 6021
Db      1833  nSerSerG1ySerSerG1ySerSerG1yG1yG1yG1ySerSerSerArgProA1aSe 1853
Qy      6022  GAGCAGCGGGGCT-----GGGGGCGAGCGCGCCGCTC 6057

```


QY 1853 rHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnG 1873
Db 6058 CCACACCC-----CACACGACCTCGCCCATCTCCCCCGGACCCAGACGCGCTTGACAGCA 6111
QY 1873 nArgProSerValLeuHisValaThrGlyMetLeuGlyTLeIleThrAlaValGluProSe 1893
Db 6112 GAGGCCAGAGTGTGTGACACACAGACATGAGGGCGCTGACCTCCGTGGAACCCGG 6171
QY 1893 rLysProThrValLeu-----ArgSerThrSerThrSerSerProValArgProAl 1910
Db 6172 CAGGCCACCGCTCTTGAGGTGGCGCAGGTCCACCTCCACCTCTGCTGCTGCCCGCACG 6231
QY 1910 aAlaThrPheProProAlaThrHisGlySerProLeuGlyGlyThrLeuAspGlyValTyrPr 1930
Db 6232 TGCCACATTCCTCCACCTGACACCACTGCCACTTGGTGGACCTTGAAAGGGCTTACC 6291
QY 1930 oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr 1950
Db 6292 TACCTCATGAGAGCCGCTCTGTTACCCAGAGACCTCTCGGGCTGCCCGCCGAGCG 6351
QY 1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG 1970
Db 6352 GGGCCGGGTGACCTGACCTGACCTCTTCTTACCAACCCCGGGCGG-----GA 6402
QY 1970 uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProAlaSerG 1990
Db 6403 GCGCGGCTCTCCACCCAGACAGAGACTCGAGCCCGCATCTTACACACCCCGACCTCAG 6462
QY 1990 yHisAlaThrTLeAlaArgThrProAlaLysValaLeuAlaProHisAlaSerProAs 2010
Db 6463 CCACACAGCCATCGCCGACACCCAGACAAAGAACTTGACACCCACATGCACTCGGGA 6522
QY 2010 pProProAlaProProAlaSerAlaSerAspProHisAlaGlnLysThrGlnSerLysPr 2030
Db 6523 CCGCGCGGGG---CCACCTCGGCTCAGATCTGCACCGAGAAAGACTCAAAATTAACC 6579
QY 2030 oPheSerTLeGlnGluLeuGlnLeuLeuArgSerLeuGlyTyrHis---GlySerSerTyrSe 2049
Db 6580 CTTTTTCATCCAGGAATTTGGAACCTCCTTCTCGGTACACAGTGGAGCTGCTACAG 6639
QY 2049 rProGluGlyValGluProValSerProValSerSerProSerLeuThrHisValAspLysG 2069
Db 6640 CCCCAGATGGGGTGGAGCCCTACAGCCCGGTGAGCTCCCAAGCTTGACCCACGACAGGG 6699
QY 2069 yLeuProLysHisLeuGlnGluLeuAspLysSerHisLeuGlnGlyValLeuArgProLy 2089
Db 6700 GCTCTCCAAACCTCTGGAAGAGCTAGAGAGAGCCACTTGAGAGGGAGCTGCGGCAAA 6759
QY 2089 sGlnProGlyProValLysLeuGlyGlyGlnAlaAlaHisLeuProHisLeuArgProLe 2109
Db 6760 GAGAGCAGGCGCCCATGAAAGCTACAGCGGGAGGCTGCCCATCTCCACATCTGGCGCACT 6819
QY 2109 uProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
Db 6820 GCCCGAGAGCAGCCCTCATGACCCACTCTCCACAGCTGCCAGGAGCTCAAAAGGTCA 6879
QY 2129 sGlnArgValAlaThrLeuAlaGlnHisTLeSerGlnValIleThrGlnAspTyrThrAr 2149
Db 6880 CCAAGAGGTGTACCTGCTGACGACATCAGAGAGTATTACGAGAGATCAACAGCG 6939
QY 2149 gHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db 6940 CACACACCGGACAGAGTGTGAGCCCTTCCCGCCCTCTACTCTCTTCCCGAGC 6999
QY 2169 aSerCyAspProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAs 2189
Db 7000 CACCTCCCTCTGCTGTGATCTTGCGCGCCACACAGAGACTCTACTCCACCCCGCA 7059
QY 2189 pHisGlyAlaProAlaArgGlySerProHisSerGlnGlyGlyLysArgSerProGluPr 2209
Db 7060 CCAATGGACCCCAAGCCCGGGGATCTCCCAACAGTGAAGGGGCAAAAGTCTCCAGAAC 7119

QY 2209 oAsnLysThrSerValLeuGlnGlyGlyGlnAspGlyIleGluProValSerProG 2229
Db 7120 CAGCAAAACATCGCTCTGGGACAGAGAGATGCCATTGAGCTGTGTCCACACAGA 7179
QY 2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspG 2249
Db 7180 GGGCATACATGAGCAGACATGCTCGAGACATCGCTTACCTCATGCTGTATCCAGACGG 7239
QY 2249 yGlnGlnThrGluProSerArgMetGlySerLysSerProGlyValaThrSerGlnProPr 2259
Db 7240 GGAACAGGGGAGGCC---AGATGGGTCTAGAGTCTTCAGGCAACACAGCCAGCGCC 7296
QY 2269 oAlaPhePheSerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysGlnG 2289
Db 7297 AACCTTCTCAGTAACTGACTGAGACCACTCCGCCATGGTGAAGTGAAGACAGGA 7356
QY 2289 uLleAsnLysLysLeuAsnThrHisValaArgAsnGluProGluTyrAsnIleSerGlnPr 2309
Db 7357 GATTCACAGAAACTCAACACCCACACCGGAACGAGCCAGAAATATATTGCGCAGCC 7416
QY 2309 oGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSe 2329
Db 7417 TGGAGCGAAATCTTCAACATGCCCGCATCTGAGACAGGCTTATGACTGTAGAAG 7476
QY 2329 rGlnAlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGlnValaIleIleArgLysAl 2349
Db 7477 CAGAGCGGTCCAGAAACACCCGACACCAACATGGGGCTAGAGGCCATTATTGAAGAGC 7536
QY 2349 aLeuMetGlyLysTyrAspGlnTrpGluGlnSerProProLeuSerAlaAsnAlaPheAs 2369
Db 7537 ACTCATGGGTAAATATCATAGTGGAGAGCCCGCCCGCTGCGCCCAATGCTTTTAA 7596
QY 2369 nProLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrLalaAspG 2388
Db 7597 CCTCTGAATGCCAGCGCCAGCTGCGCTGCTGCTATGCCATTAACCACTGTGACGG 7656
QY 2388 yArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyAr 2408
Db 7657 AGGAGTGACCAAGCACTCACTCGCAGGTGAGGTGGAAAGCCAAAGTCTCTGGCAG 7716
QY 2408 gProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProPr 2428
Db 7717 ACTTAGAGCGCAAAAGCCAAAGTGGCAGCAGCAGGCTGACCTCCGAGAACCGACCCC 7776
QY 2428 oSerValSerSerValHisSerGlnGlyAspCyAsnArgArgThrProLeuThrAsnAr 2448
Db 7777 TTCTGTCTCTCACTCACTCAGAGGGAGACTGCATTCGCGAACACCACTCACCAACCG 7836
QY 2448 gValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuI 2468
Db 7837 TGTGTGGAGAGACGGGCCCTCATCTGACAGGTCCACGCCATTCCCTTACAACTTTGAT 7896
QY 2468 eMetArgLeuGlnAlaGlyValaMetAlaSerProProProProGlyLeuProAlaGlySe 2488
Db 7897 TATGAGGCTTACAGGAGGTGTCTATGCTCCCGCCCGCCACTGCGCTTGCGACAGGAG 7956
QY 2488 rGlyProLeuAlaGlyProHisHisAlaTrpAspGlnGluProLysProLeuLeuCySe 2508
Db 7957 CGGGCCCTTAGCTGTGTCCCAACAGCGCTGGGATGAGAGGCCAAGCCATGCTGTGTTC 8016
QY 2508 rGlnTyrGluThrLeuSerAspSerGlu 2517
Db 8017 ACAGTATGAGACCTCTCGACAGCGAG 8044

RESULT 5
AF125671 8544 bp mRNA linear ROD 04-Apr-1999
LOCUS AF125671
DEFINITION Mus musculus silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform (Smrte) mRNA, complete cds.
ACCESSION AF125671 GI:4559295
VERSION AF125671.1
KEYWORDS Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 8544)
AUTHORS Park, E.-J., Schreier, D. J., Yang, M., Li, H., Li, L. and Chen, J. D.
TITLE SMRT, a silencing mediator for retinoid and thyroid hormone
receptor corepressor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
MEDLINE 99199215
PUBMED 10097068
REFERENCE 2 (bases 1 to 8544)
AUTHORS Chen, J. D.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology,
University of Massachusetts Medical School, 55 Lake Avenue North,
Worcester, MA 01655, USA
FEATURES
source
1. 8544
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1_8544
/gene="Smrte"
160..7548
/gene="Smrte"
/function="transcriptional corepressor"
/note="SMRTE"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform"
/protein_id="A022972.1"
/db_xref="GI:4559296"
/translation="MSGSTQPYAQWRAAPRYPHGISYVQIARSHTDVGLLEYOH
HRDYSHTLSPGSIITOPORRPSLSRSGSRSGEILRLRPSRFLRFLRICKDIEF
TSSKPRLELEPTTLRPSPLATGPPSSSDITKRSILAKGIEFPPSPRPADDEL
ELAPSLKSEBELQNLDRVDREITWEEQISLKKRQQLLEAKPEPEKVPSP
PIESKRSLVQIITYDENRKAENAHRLBGLGQVELPYNQSIRQYHENTKIQA
MRKLLYFQRNHAARKWQRCORYDOLMEAMEKVRLENNPRRAKESKVRBY
EKQPEIRKORELOERMOSRGVSGSLMSARSEHVESEIDGLSEONLEKQRO
LAVIRHVVRDOORIKETINMGMDMPVYKDRQYTNMSEDERPTKMPQHPN
FGILASFLEKRTVAECVLYYLTKKMENTKSLVRSYRRKGSQ0000000000000000
AASSQSEKEKEKEADNEEKQADNEKEBELSKERTDISENDEKAVASKK
TANSQGRKGRITRSMANANHEATAPPOSSSLAEMNEMSRLEBEMETAKKLT
EHAANWSALAIAMVGSKTVSQCKNIFYNKKRQNLDEILQOKLMEKERNARKKKT
PAAASETAPPAABDEEMASGASANEELAEASGASGNEVAVGSGSPAAVN
NSDTRSPSPSEATKDGPKPTGTEALPAATOPVPPPEPAAAPAPSPADSG
PPSPSPAPAPATVDKQAPAPAPAPQTDAKGQKSEAEITVGRPEPABEP
PEPSVSDHKEETEEPEDEKAGTEALITVSEAPLVEEASFAATVTKSSGATODS
DSATCSADBEVDEEGDKGRLSPRPSLITPAGDPASTSPQPLDLQKQRAAI
PPIVTVKHPREDIVPKPVPVPPTQHLQPEGDVSOQSGSPGKRSVPAPAK
BAKAPAPAPFPGQSYRLSPHAGRLSPHREVIKTSTRADPLFSTYPPPLPL
GLHDSARPLPPIPISNPPLISAKHNGEOLGATISGAGVOLRVHSEAKKMG
PLTMEPLAVDPKKTALAPPEVASPRASQYPCRRQLOLYHPRPADVLKGT
ISRIVEGDSPLDKARBDTLPRGHVIEBKGAHVSIEBGSVSCSDKDGSSGP
PHETAPKRTYDMRGRVGTIVTSASIEBGMRAIPEQSPHLKEOHILRGSLTQGI
RSYVEAQDYILREBAKILKREGTPPPPPREDLETYKRPRLDLKPKPHKEV
ATYKAGRSIHIEPRELRTPEPLAPAPRLKEGSLTQGPRLYDGAEPKPEV
BSITGSGRPALPLIDIMADALERACVEESLKSRSSTSGAGSITRGAIVVVP
ELKPRQSPITTYEDHAPRTSHLPRGSPVTRTREPRLQEGSLSKASQDRULTVP
RETAQPSHSTVEPHHPISVYHLRGVGVLYRGHPLADPSTIPGRLIEAAA
AAVYLPRHLAPSPTYPLHYPLIRIGRPDPALENQTLINDYITSQOMHNHNSAMA
QRADMLRGLSPRESSILANSAGRGIIDLSQVPHLPLVLPFGPAPATIDILAVLP
TAPPPSRHSSSPLSPGPTHLAKPATSSSERREBEREKSLITSTTYEHAPI
WRCTGSSGAGGSSRPASHTQHSPIPTDOLQORBEVLNMTSGKGVNSVEGT
PYLARSSTSSVPYRAPATPPPATPCPLGSLGCVPTIMEPVLLPETSIVARPEPR
VDDGHAFLTKPAREPASSPSSSEPSLAPSSSHALAKTAKSLAPHASAPDPG
PTASDLHREKTSKPFSIOELRLRSYGYSAGYSDVEYSPVSPBLTDKGLS
KLELEKSHIEGELRHQKQPMKLSAEANHLPHLRPLPSQSSSPLQIATAGIGH
QRYVTLAHOHISEVITODYTHHPOOLSGPLPAPLYSPGASCVLLDRSPDLIYRP
PDGCTPARGSPHSEGRSPESKTSKESDAITPVSPBEMTREGHARSVYVLL
YRDGEGEPHMSKSPONTSQAPFSPKLTESANVYSKKQIKINKLTHNNEBYL
NIGQPTETLPMNPAITGAGLMTCRSAQVOEHASTNMGLEAIIRKALMGKWEPPPP

ORIGIN
Alignment Scores:
Pred. No.: 1,5e-210 Length: 8544
Score: 10832.50 Matches: 2149
Percent Similarity: 87.94% Conservative: 90
Best Local Similarity: 84.41% Mismatches: 197
Query Match: 81.97% Indels: 117
DB: 10 Gaps: 33
US-09-522-753-5 (1-2517) x AFI25671 (1-8544)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyPro 20
DB 160 ATGTCAAGATCCACACAGCCTGTGCGACAGACATGCGGGCTGTGAGCCCGCTACCCA 219
QY 21 ProHisSerLeuSerTyProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 220 CCCATGCGCATCTCCACCCGGTGCAGATAGCCGGTCCACACGACGATGGGGCTGCTT 279
QY 41 GluTyrglnHisHisSerArgAspTyraIAserHisLeuSerProGlySerIleIleGln 60
DB 280 GAGTACCAACACACACCCCGTGACTACACCTACACCTTACACCGGTTCCATCATCCAG 339
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValangLuarGserGln 80
DB 340 CCACAGAGAGAGCGGCGCTCATCGCTGCAGATTCCACCTCGGAGTGACAGGCTCAG 399
QY 81 GluLeuHisLeuArgProGluSerHisSerTyLeuProGluLeuGlyLysSerGluMet 100
DB 400 GAGCTTCACCTCGCGCCTCAGTCCCGCACGCTGCTCCGAGAGCTGGGCAAGCCGACATA 459
QY 101 GluPheIleGluSerTyArgProArgLeuGluLeuProAspProLeuLeuArgPro 120
DB 460 GAATTACCGAGAGCAAGGCGCCCGCTGAGACTACACCGATACCTGCTGCGGCCA 519
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 520 TCACCCCTCTGCGCACCTGGGACGAGAGTGGGTGGAAGACTTACCAAGACCGTAGC 579
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 580 CTGGCAGGCAAGCTGAGCGCTGTGTCACCTCCAGTCCCGGACCGCTGAGCTA 639
QY 161 GluLeuValProProArgLeuSerLysGluLeuIleGlnAsn--MetAspArgVal 179
DB 640 GAGCTGGCGCATCTTGACTGTCCAAAGAGACTGATCCAGAACAGATTGACCGCGTG 639
QY 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGln 199
DB 700 GACCGGAGATCACCATGTGAGAGCAGCATCTCCAGACTGAAGAAAGACGACAG 759
QY 200 LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProIle 219
DB 760 TTGGAGAGAGAGGCGCCCAAGCGCCGGAACCGGAACCTGTGTGCGCACACCCATA 819
QY 220 GluSerLysHisArgSerLeuValGlnIleIleTyAspGluAsnArgLysLysAlaGlu 239
DB 820 GAATCAAGACACGAAGCTGGTGCAGATCATCTTACAGTGAAGAACCGGAAGAACCGGA 879
QY 240 AlaAlaHisArgIleLeuGluGluLysGluProGlnValGluLeuProLeuTyraGln 259
DB 880 GCGGACACCGGATCTTGAAGAGCGTGGGGCCCAAGTGAAGCTGCTGTGACACAG 939
QY 260 ProSerAspThrArgGlnTyHisGluAsnIleLysIleGlnGlnAlaMetLargLys 279
DB 940 CCGTCTGACACACGCGCAGTACCATGAAAAATCAATCAATCAATCAATCAATCAATCAAT 999
QY 280 LeuIleLeuTyPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299

Dh 1000 CTGATCTTGTATCTTTAAAGCGGAGAAACACGCGCGCAAGACGTGGGAAACAGCCTTCTGC 1059
Qy 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluTyrValGluArgIleGluAsnAsp 319
Dh 1060 CAGCGCTATGACCAAGCTCATATGAGCGCGTGGAGAAAGAGTGAAGCCGATGAGAACAT 1119
Qy 320 ProArgArgArgAlaValGluSerIleValArgIleTyrGluTyrGluPheProGlu 339
Dh 1120 CCGGAAAGAGGGCCCAAGAGACAGAGGTGAGGAGTACTACGAGAAACATGCTCCGGAG 1179
Qy 340 IleArgIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 359
Dh 1180 ATCCGACAGACGCGGAGCTGACAGAGCATGACAGACGCGGCGCACGCTGGCGCT 1239
Qy 360 GlyLeuSerMetSerAlaAlaArgSerGluIleGluValSerGluIleIleAspGluLeu 379
Dh 1240 GGGCTTCCATGTGGCTGCGCGCGAGTGAAGCATGAGGTTTCTGAGATCATTTGAGGCTTG 1299
Qy 380 SerGluGlnGluAsnLeuGluTyrGluGluMetArgGlnLeuAlaValIleProMetLeu 399
Dh 1300 TCTGACAGAGGAACCTGAGAACAGATGCGCCAGCTGCGCGTGTATCC--GCCATGTTC 1357
Qy 400 TyrAspAlaAspGlnGlnArgIleIlePheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Dh 1358 TACGACCC-GACCAAGACAGATCAAGTTCATCAATGATGACTCATGATGATGACCCC 1416
Qy 420 MetIleValTyrIleAspArgGluValIleMetAsnMetTyrSerGluGluGluGluThr 439
Dh 1417 ATGAGGCTTACAGAGACGCTGAGTTACCAACATGTGAGCGAGACAGAGAGGACACC 1476
Qy 440 PheArgIleGluPheMetGlnIleAspIleAsnMetGluLeuIleAlaSerPheGluGlu 459
Dh 1477 TTCGGTAGAAGTTATGACCACTTAAAGACTTTGGCTGATTTGCTCTCTCTGAG 1536
Qy 460 ArgIleThrValAlaGluCysValLeuTyrTyrTyrLeuThrIleValAsnGluAsnTyr 479
Dh 1537 AGAAGACGCTGCTGAGTGTCTCTTACTTACTGACCAAGAAAGAAATGAAATTAAC 1596
Qy 480 LysSerLeuValArgArgSerTyrArgArgArgIleLysSerGlnGlnGlnGlnGln 499
Dh 1597 AAGAGCTTGGTAGGCGGAGTATGCGCGCGTGGCAAGACCCAGCAGCAGCAGCAGCAG 1656
Qy 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Dh 1657 CAACAACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1704
Qy 520 LysArgGluTyrGluTyrGluTyrGluTyrGluTyrGluTyrGluTyrGluTyrGlu 539
Dh 1705 AAGGAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGG 1764
Qy 540 AsnAspIleGluAspIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 559
Dh 1765 AACGAGAGAGAAAGAACTCAGCAAGAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1824
Qy 560 GluIleGluAlaValAlaSerIleValArgIleThrAlaAsnSerGlnGluArgIle 579
Dh 1825 GAGAAAGAGCCCTGCGCTCCAAAGCCGCAAACTCCCAACAGCCCAAGCCCGCGCAA 1884
Qy 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Dh 1885 GGGCGGTATCAGCGCTCATGTGCCAAAGAGCCCAATATGAGAGAGACAGACCCCAAG 1944
Qy 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTyrPheGluGluGlu 619
Dh 1945 CAAGGTCAGAGCTGCGCTTCCATGAGATGAACAGAGTTCCTGCTGACTGAGAGAGAG 2004
Qy 620 MetGluThrAlaValIleValIleValIleValIleValIleValIleValIleValIle 639
Dh 2005 ATGAGAGACAGCAAAAGAGCCCTCTCGAAATGCGAGAGAACTGTGACGCAATGGCCGC 2064
Qy 640 MetValGluSerIleThrValSerGlnCysLeuAsnMetIlePheAsnTyrIleValArg 659
Dh 2065 ATGGTGGCTCCAAAGACCGGTGCTCCAGTGTAAAGACTTCACTCAAGAAAGAGG 2124

Qy 660 GlnAsnLeuAspGluIleLeuGlnGlnIleIleLysLeuIleMetGluTyrGluValArgAsnAla 679
Dh 2125 CAGAACTCGAGCGAAATCTTCTACAGACACCAAGCTTAAAGATGAGAGAGAGAGAGAGCT 2184
Qy 680 ArgArgIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 699
Dh 2185 CGGAGAGAAAGAAAG 2244
Qy 700 ValGluAspGluGluMetGluAlaSerGluValSerGluValSerGluValSerGluVal 719
Dh 2245 GCTGAGAGACGAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 2304
Qy 720 GluAlaGluAlaLeuValAlaSerGluValAsnGluValProArg---GlyGluCysSerGly 738
Dh 2305 GAGGACAAAGCTTACAGAGCTCTGAGAAATGAGTTCCCAAGATTTGGAGATGCAATGAGC 2364
Qy 739 ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProIleThrGluAla 758
Dh 2365 CAGCTGCTGTCAACCAAGCTGTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2424
Qy 759 AlaIleAspThrGluGlnAsnGlyProLysProProAlaThr-----LeuGluValAsp 776
Dh 2425 ACGAAGGACACT-----GGGCTTAAACCCACTGCACTGAAGCATGCGCGCTGCC 2475
Qy 777 GlyProProGluProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Dh 2476 ACCGACGACCTGTCTCT-----CTTCCAGAGAACCCGACAGAGCCCTGCTGAGACC 2529
Qy 797 ThrProAlaSerGluAlaThrGluValAlaProThrProProProAlaProProSerProSer 816
Dh 2530 TCCCACTGCTGATGCGCAGAGGCCACCAATCCCAAGAGCTTCC---CATCATCTGACC 2586
Qy 817 AlaProProProValValIleProLysGluGluTyrGluGluGluGluGluGluGluGlu 836
Dh 2587 GCACCCCGGCTACTGTGAGCAAGAGTAAACAAGAACCCCGGCTGTCTCCAGCTCCAG 2646
Qy 837 ValGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 856
Dh 2647 ACGAGAGATCCAAAGAGACGAGAGTGTGAGGCGCAGAG-----ATCGATGTGGGA 2697
Qy 857 LysAlaGluGluPro-----ValIleSerGluCysThr 867
Dh 2698 AAGCAGAGAGAGCCGAGGCTGTGAGAGACCCCGGAGAGTGTAAAGATGACCAAG 2757
Qy 868 GluGluAlaGluGluGluProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr 886
Dh 2758 GAGGAGACCGAGAGAGAGCTGAAAGCAAGGCAAGGCGCAGAGGCCATTGAAACTGTG 2817
Qy 887 AlaGluGluAlaLeuLysAlaGluLysLysGluGluGluGluGluGluGluGluGluGlu 905
Dh 2818 TCTGAGCACCACTTAAAGGTGAG-----GAGGCTGTGAGCAAGGACGCTGTGACCAAG 2871
Qy 906 AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Dh 2872 GGTTCACAGCTCAAGTGCACCCAGAGACAGTCACTCCAGTGCACCTGCAATGCCAGTAG 2931
Qy 926 ValAspGluAlaGluGluGluLysAspLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
Dh 2932 GTGAGCAACCCCAAGAGAGGTGCAAGGCGAGCTGTGTACCAAGAGCCAGCTTCTTC 2991
Qy 946 ThrProThrGlyAspProArgAlaAlaAlaSerProGlnLysProLeuAspLeuValGln 965
Dh 2992 ACCCGGCTGGAATCCCGGCGCAAGTCACTTCCGCCAAGAGCGCTGAGACCTGAAACAG 3051
Qy 966 LeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValIleGluProPro 985
Dh 3052 CTGAAGAGAGAGACAGCGCCCATCCCCCTATC---GTCAACCAAGGTCCATGAGCCCCC 3108
Qy 986 ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeu 1005
Dh 3109 CGGAGAGACACAGTACCCCAAGGCAAGTTCCTCCCTGTGCTTCCACCAAGACGACCTTA 3168

QY 1006 GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgIleuYlySerArgSer 1025
 DB 3169 CAGCCAGAGGGTGCCTGTCCTCAGAGTCGGAGAGAAAGTCCAGTGGCAAGTCCCGCAGC 3228
 QY 1026 ProAlaProProAlaAspIlySerGlu-----AlaPheAlaAla 1037
 DB 3229 CCAAGTCCCTCTGGCCAGAAAGGAGAGAAACCCGCATTTCTTCCGCTTTCCCAACT 3288
 QY 1038 GluAlaGlnIlyLeuProGlyAspProGlyAspTrpThrSerGlyLeuProPheProVal 1057
 DB 3289 GAGGGCCA-AGGTACCGACTGAGGCCCCACGCTGGTCAATCGCT---GCCTTCCCAT-- 3342
 QY 1058 ProProArgGluValIleuYlyAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
 DB 3343 CTTCCACGGAGGGATCAAGACTTCCACACGGGCT---GACCTT---CTCTTCTCTAC 3396
 QY 1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValIleuPro 1097
 DB 3397 ACAACCCCGGTACACCGGCTCTGGGCTTCACAGATAGTCCGGCCGCTCGCA 3456
 QY 1098 ArgProProThrIleSerAspProProProLeuIleSerSerAlaIlyHisProSerVal 1117
 DB 3457 CGTCCCCC---ATCTCTAACCCCCCAACCCCTCATCTCTCTCCAAAGCATCCCGCGTA 3513
 QY 1118 LeuGluArgGlnIleGlyAlaIleSerGlnIlyMetSerValGlnLeuHisValProTyr 1137
 DB 3514 CTTGAGAGGAGCGATGGGTGCATCTCCAGAGGATGTCAAGCTTCGTGGCTGCAC 3573
 QY 1138 SerGlnHisAlaIlyAlaProValArgProValThiMetGlyLeuProLeuProMetAsp 1157
 DB 3574 TCGAGCATGTGCCAAG---CCCATGGGCGCTCTCACATGAGGTGCGCTTGGCGTGGAC 3630
 QY 1158 ProIlyLeuAlaProPheSerGlyValIlyGlnGlnIleuSerProArgGlyGln 1177
 DB 3631 CCTAAGAGCTG----- 3642
 QY 1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValIleuArgGly 1197
 DB 3643 -----GGG 3645
 QY 1198 ThrAlaLeuGlySerValProGlyGlySerIleThrIlyGlyIleProSerThrArgVal 1217
 DB 3646 ACAGCACT-GGCTCCGCCACAGTGAAGCATCACAAAGGCGCT-CCCACTACCCGCGCT 3703
 QY 1218 ProSerAspSerAlaIleThrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
 DB 3704 GCAGAGGCGCC-----ACCTACAGAGCTCTATCACCCACC-ACGCCGAGACGTC 3756
 QY 1238 LeuTyrIlySerIlyThrIleThrArgIleIleGlyIleuAspSerProSerArgIleuAspArg 1257
 DB 3757 CTCTACAAAGGTATCATACAGAGATCGTGGTGAAGACAGCCCAAGTGGCTTTGACCGG 3816
 QY 1258 GlyArgGluAspSerIleuProIlyGlyHisValIleTyrGlnGlyIlyIlySerGlyHisVal 1277
 DB 3817 GCACGAGAGACACCTGCCCAAGGGCGCATGTATGAGGCAAGAAAGGCCACGTC 3876
 QY 1278 LeuSerTyrGlnIlyGlyMetSerValThiGlnCysSerIlySerGluAspArgIlyArgSerSer 1297
 DB 3877 CTATCTATGAAGGTGATATGTCGATGTCAAGTCTCTTAAGAGATGAGAGAGACAGC 3936
 QY 1298 SerGlyProProHisGlnThrAlaAlaProIlySerArgThrTyrAspMetMetGlnIlyArg 1317
 DB 3937 TCGGGCCCAACCCATGAGACTGCCGCCCTTAACGCACTTATGATCATGAGAGGGCGGT 3996
 QY 1318 ValGlyArgAlaIleSerSerAlaSerIleGlnGlyLeuMetGlyArgAlaIleProPro 1337
 DB 3997 GTAGGACGAGCTGTCACTCAGCCAGCATGAGGAGACTCATGGCGCGCGCATC---CCT 4053
 QY 1338 GluArgHisSerProHisIleuIlyGlnGlnHisIleArgIlySerIleThiGln 1357
 DB 4054 GACAGCAAGCGCC---CACCTCAAGAGACAGATCATCCAGAGCTCATACAGCAA 4110
 QY 1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgIlyAlaIlyLeu 1377

DB 4111 GGATCTCCCAAGGTCTCTATGTGAGGCGAGAGACTTATCGGGAGGCGCAAGCTC 4170
 QY 1378 LeuIlyArgGlnGlyThiProProProProProSerArgAspLeuThrGluAlaTyr 1397
 DB 4171 TTGAAGCGAAAGGAGACCACTCCCCACCAACCTCGGGAGACTGATGAGACTTAC 4230
 QY 1398 IyethrGln-----AlaLeuGlyProLeuIlyLeuIlyProAlaHisGlnGlyLeu 1414
 DB 4231 AAGCCCGGCGCTGGACCTCTGGGTCCTCGAAGCTGAACCCGACTCACAGAGGTGTG 4290
 QY 1415 ValAlaThrValIlyGlnIlyAlaIlyArgSerIleHisGlnIleProArgIleGlnIleuArg 1434
 DB 4291 GTAGCAACTGTAAAGAGGCGGCGCTGTATCCAGAAATCCGAGAGAGAGCTGCGC 4350
 QY 1435 HisThrProGluIleuProLeuAlaProArgProLeuIlyGlnIlySerIleThiGlnGly 1454
 DB 4351 CGCACCTGAGTACCTCTGGCACCAAGGCTCTGAAGAGGTTCCATCACCCAGGCG 4410
 QY 1455 ThrProLeuIlyTyrAspThrGlyAlaSerThrThiGlySerIlyIlyHisAspValArg 1474
 DB 4411 ACCCACTCAAGTACGACTGTGGGGCACCTCCACTGGCACCAAGAAACACGCGCGCC 4470
 QY 1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
 DB 4471 TCCATCATGGAGAGCGCCGCGCGCTTCTCCCTGCCCTGCACCCGCTGACATATGGCT 4530
 QY 1495 AspAlaArgAlaLeuGlnIlyArgIlyCysTyrGlnGlnIlySerLeuIlySerArgProGlyThr 1514
 DB 4531 GACGCCCGGACCTGAGGCTGCTCTGTAAAGAGTCTGAAGACCGGTCAGGAGAC 4590
 QY 1515 AlaSerSerGlyIlySerIleAlaArgIlyAlaProValIleValProGluLeuGly 1534
 DB 4591 AGCAGTGTGACAGGGGCTCCATCACAGTGGGGCTCCAGTGCCTGCTGAACTGGGC 4650
 QY 1535 IyProArgGlnSerProLeuThrTyrGlnAspHisGlyValProPheAlaGlnHisLeu 1554
 DB 4651 AAGCCACGGCAAGCCCACTGACTTAAGAAACACAGGGGCGCCCTTCAACGACCTG 4710
 QY 1555 ProArgGlySerProValThiMetArgGluProThrProArgLeuGlnIlySerLeu 1574
 DB 4711 CACGTGGCTCCCTGTGACACAGAGAGCCACCGCCACGCTTCAAGAAAGCACCTC 4770
 QY 1575 SerSerSerIlyAlaSerGlnAspArgIlyLeuThrSerThrProArgGluIleAlaIly 1594
 DB 4771 CTATCCAGCAAGCGTCCAGAGACCGAAGTGAATCACTAACCCCGGAGATCGCCAA 4830
 QY 1595 SerProHisSerThrValProGlnHisIleProHisIleProIleSerProTyrGlnHisLeu 1614
 DB 4831 TCCCAACAGACGCTGTGCCGAGCACACCTCACCCCATCTCCCTATGAGCACTTG 4890
 QY 1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
 DB 4891 CTCGGGAGGTGACTGTGTGAACCTGTACCGTGTCAATCCCATTTGGCTTTGACCCC 4950
 QY 1635 ThrSerIleProArgIlyIleProLeuAsp-----AlaAlaAlaAlaTyrTyrIleuProArg 1653
 DB 4951 ACCTCATACCCCGAAGATCCCTCTGAAGACGAGCGGACGCTTAACCTGCGCCCG 5010
 QY 1654 HisLeuAlaProAspProThrTyrProHisIleuTyrProProTyrIleuIleArgIlyTyr 1673
 DB 5011 CACTTGGCCCGCAGCCCACTAACCCACGCTGTACCACTTACCTCATCCGCGGCTAC 5070
 QY 1674 ProAspThrAlaAlaLeuGlnAspArgIlnThrIleIleAspAspTyrIleThrSerGln 1693
 DB 5071 CTTGACAGCGCGCTCTGAGAAACCGCACCATATATCAATACATCACTGCGAG 5130
 QY 1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgIlyLeu 1713
 DB 5131 CAGATGACACAAAGCTGCTCCGCAATGGCCAGCGTGTGACATGTGTGAGGGGTCTG 5190
 QY 1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733

Dh 5191 TCACCGGAGAGCTCTCGCTGGCCCTCATTTATTCGCTGGCCCAAGAGCATTTACGAC 5250
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Dh 5251 CTGCCCCAAGTGCACACCTGCCCTGTGTGTGCACCAAGGACCAAGGACCCCTGGCACCC 5310
Qy 1754 AlawerAparGluLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Dh 5311 GCCATCGACCGCCTTGCTACTCTCCCACTGCGCCCACTTCAGAGAGCGGCACAGT 5370
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793
Dh 5371 AGCTCAGCGCTGTCTCCAGAGAGGCCCACTCACTTAACCACTGCGCACATCTTCA 5430
Qy 1794 SerGluArgGluArgAparAparGluArgAparAparGluArgGluLysSer 1813
Dh 5431 TCGGAGGGGAGACGGAGACGTGACGGGAGACGAC-----AAGTCC 5472
Qy 1814 IleuThrSerThrThrThrValGluHisAlaProIleTpaArgProGlyThrGluGln 1833
Dh 5473 ATCTCAGCTACACTACAGTACATGACATGACCCATCTGAGACCTGTGTAGAGAGCAG 5532
Qy 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Dh 5533 AGCAGCGGGGCT-----GGGGGACAGCAGCCCGCCCTCC 5568
Qy 1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAparAlaLeuGln 1873
Dh 5569 CACAC-----CACGACCTGCCCCCTCCCGGACCCAGAGGCGCTTGCACAG 5622
Qy 1874 ArgProSerValLeuHisAenThrGlyMetLysGlyLysIleThrAlaValGluProSer 1893
Dh 5623 AGGCCAGTGTGTGCACACAGAGCATGAGGGCGTGTCACTCGTGGAAACCGGCG 5682
Qy 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
Dh 5683 ACCGCCACGGTCTGAGTGCACCTCCACTTCGCTGCGCCAGCTGCGACATTC 5742
Qy 1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAparGlyValTyrProThrMet 1933
Dh 5743 CCACCTGCCACCACTGCCCACTTGTGGACCTTGAAGGGGTCTACCTACCTCATG 5802
Qy 1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAla 1953
Dh 5803 GAGCCGCTCTTACCAAGAGACCTTCGGGTGCGCGCCGAGCGGCCCGGTG 5862
Qy 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Dh 5863 GAGGGTGGCATCTTCTCCACCAACCCCGCGCGG-----GAGCCGCTCC 5913
Qy 1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Dh 5914 TCACCCAGCAAGAGCTCCGAGCCCGATCCCTAGACACCCCAAGCTTCAGCCACACAGCC 5973
Qy 1994 IleAlaArgThrProAlaLysAenLeuAlaProHisHisAlaSerProAparProAla 2013
Dh 5974 ATGCGCCGAGCCCGAGAAAGAGCTTGCACCCCACTGCAATGCCAGCCGCGCGGG 6033
Qy 2014 ProProAlaSerLysAparProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Dh 6034 ---CCCACTCGGCTCAGATCTGCACCGAGAAAGCTCAAGTAACCTTTTCATC 6090
Qy 2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerLysSerProGluGly 2052
Dh 6091 CAGGAATTGAATCTCGTTCTCTGGGTACCAAGTGAAGCTGTGCTACACCCCGATGG 6150
Qy 2053 ValGluProValSerProValSerSerProSerLeuThrHisAparGlyLeuProLys 2072
Dh 6151 GTGGAACCCATCAGCCGGGTGAGCTCCCGAGCTGACCAAGCAAGGGGCTCTCCAA 6210
Qy 2073 HisLeuGluGluLeuAparLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Dh 6211 CTTCTGAAGAGCTTGAAGAGCCACTTGGAGGGGAGCTGCGGACCAAGCAGCAGGC 6270

Qy 2093 ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAparProLeuProGlySer 2112
Dh 6271 CCATGAAGCTCAGCGGGAGAGGCTGCGCCATCTCCCACTGTGGCGGCACTGCCAGAGC 6330
Qy 2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Dh 6331 CAGCCCTCATTCAGCCCACTCTTCCAGACTGCCCCAGGATCAAAAGGTCAACAGAGGTG 6390
Qy 2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAparLysThrArgHisHisPro 2152
Dh 6391 GTCACTTGCTCAGCAATCAGAGGTCAATTAAGAGACTACACCCGACACACCG 6450
Qy 2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Dh 6451 CAGCACTCAGTGGCCCCCTTCCGCCCTCTACTCTCTTTCCGAGCCAGCTGCGCT 6510
Qy 2173 ValLeuAparLeuArgArgProProSerAparLeuTyrLeuProProProAparHisGlyAla 2192
Dh 6511 GTGCTGATCTTGCAGCGCCCAAGTGAAGCTTACTCCCAACCCCGACATGGAC 6570
Qy 2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAenLysThr 2212
Dh 6571 CAGCCCGGGGATCCCCCACTGAGAGGGGGCAAAAGTCCCCAGAAACCAAGCAACA 6630
Qy 2213 SerValLeuGlyGlyGlyAparGlyIleGluProValSerProProGlyLysMetThr 2232
Dh 6631 TCGGTCTGGGACAGAGTGAAGATGCCATTGAGCTGTGTGCCCAAGAGGCGATACT 6690
Qy 2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAparGlyGluGlnThr 2252
Dh 6691 GAGCCAGCATCTCGAGCGCTGTGTACCACTGTGTATGAGACGGGAAACAGGCG 6750
Qy 2253 GluProSerArgMetGlySerLysSerProGlyAenThrSerGlnProProAlaPhePhe 2272
Dh 6751 GAGCCC---AGATGGGCTCTTAAGTCTCCAGGAAACCAAGCAGCGCGCTTCTTC 6807
Qy 2273 SerLysLeuThrGlnSerAenSerAlaMetValLysSerLysLysGlnIleAenLys 2292
Dh 6808 AGTAAGTCACTAGAGCAATCCGCCATGTGAAGTCAAGAAAGCAGAGATCAACAG 6867
Qy 2293 LysLeuAenThrHisAenArgAenGluProGluTyrAenLysSerGlnProGlyThrGlu 2312
Dh 6868 AAACCTCAACCCCAACCGGAAAGAGCCAGATTAATATGCGAGCTGGGACGAA 6927
Qy 2313 IlePheAenMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Dh 6928 ATCTTCAACATGCGCCCACTCACTGAGCAGGCTTATGACCTGTGAAGCCAGCGGTG 6987
Qy 2333 GlnGluHisAlaSerThrAenMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Dh 6988 CAGAAACAGCCCAACCAATGGGCTGAGAGCCATTTAGAAAGGCACTCATGGGT 7047
Qy 2353 LysTyrAparGlnTpaGluGluSerProProLeuSerAlaAenAlaPheAenProLeuAen 2372
Dh 7048 AAATATGATCAGTGGGAAAGCCCGCGCTGCGGCGCAATGCTTTTAACTCTGAAT 7107
Qy 2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAparGlyArgSerApar 2391
Dh 7108 GCCAGGCGCAAGTGTGCCCTGTCTATGCCCATAACCACTGCTAAGCAGAGAGTAC 7167
Qy 2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Dh 7168 CAGCACTCACTCGCAGAGTGAAGTGGAAAGCCAAAGTCTCTGGACAGCTTACAGC 7227
Qy 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAparArgProProSerValSer 2431
Dh 7228 CGAAAGCCAAAGTGCACACCAAGGCTTAGGCTCGGAGACCGACCCCTTGTCTCC 7287
Qy 2432 SerValHisSerGlnGlyAparCysAenArgArgThrProLeuThrAenArgValTpaGlu 2451
Dh 7288 TCAGTCACTCAGAGGGGAGCTGCAATCGCCCAACCACTCAACCAAGCTGTGTGGAG 7347

Db 467 ----- 467
 Qy 201 GluGluGluAlaAlaIleValProProGluProGluValSerProProIleGlu 220
 Db 467 ----- 467
 Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
 Db 467 ----- 467
 Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
 Db 468 ----- 485
 Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleLeuGlnAlaMetArgLysLeu 280
 Db 486 TCTACACACCCCGATACATGAAAACATCAAAATTAACCGGCGATCGGAAAGAGCTG 545
 Qy 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
 Db 546 ATCTTGACTTAAAGCGAGAACCAAGCGCGCAAGAGTGGAACAGCGCTTCTGCGAG 605
 Qy 301 ArgTyrArgGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
 Db 606 CGCTATACACGCTCATGAGCGCTGGAGAAAGATGAGCCCATGAGAACATCCG 665
 Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
 Db 666 CGAAGAGGCGCCAGAGAGACAGTAGAGATCTACGAAACAGTTCCCGAGATC 725
 Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgLysSerGly 360
 Db 726 CGCAGACAGCGGAGCTGCAGAGACGCGATGCGAGCGAGCGGTGGCCAGCGTGGAGTGG 785
 Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
 Db 786 CTCCTCCATGTCGCTGCGCCGAGTGAACATGAGCTTCTGAGATCATTTGAGCTTGTCT 845
 Qy 381 GluGlnGluAsnLeuGluLysGluMetArgGlnLeuAlaValIleProProMetLeuTyr 400
 Db 846 GAGCAGAGAACTCGAGAAAGACAGATGCGCAGCTGCGCTGATCCGCGCATTTGTAC 905
 Qy 401 AsnAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
 Db 906 GAGCGGACCGAGAGATCAAGTTCATCAACATGATGACCTCATGATGAGCCCATG 965
 Qy 421 LysValIleTyrLysAspArgGlnValMetAsnMetLysSerGluGlnLysGluThrPhe 440
 Db 966 AAGGTCTACAAAGACCTGTGAGTTACCAACATGTGAGCGAGCGAGAGGACACCTTC 1025
 Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
 Db 1026 CGTGAGAAATTATGACGACCCCTAAGAACTTTGGCTGATGTGCTCATTTCTGGAAGA 1085
 Qy 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
 Db 1086 AAGACGCTGCTGATGTCTCTCTTACTACTGACCAAGAAATGAAAATTTACAG 1145
 Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
 Db 1146 AGCTTGTTGAGCGGAGCTATCGCGCGCTGCGCAAGCGACGACGACGACGACCAA 1205
 Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys 520
 Db 1206 CAACGACGACGACGACGACG-----ATGCGACGAGACGACCGACGAGAGAG 1253
 Qy 521 AsnGluLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysProGluValGluAsn 540
 Db 1254 GAGGAGAAAGAGAGAGAGAGAGCGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
 Qy 541 AsnLysGlnLysLeuLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 560
 Db 1314 GAGAAAGAAAGAACTCAGCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373

Qy 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleArgArgLysGly 580
 Db 1374 AAAAGAGCCCTGGGCTCTCAAAAGCCGCAAACTCCCAAGACCAAGCCCGCCAGAGCC 1433
 Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 600
 Db 1434 CGTATCAGCGCTCATGCGCAAGAGCCCAACCATGAGAGAGAGAGAGAGAGAGAGAG 1493
 Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
 Db 1494 AGTTCAAGCTGCTTCATGAGATGACGAGAGTTCTGCTGAGCTGAGAGAGAGATG 1553
 Qy 621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTyrPheAsnTyrLysArgGln 640
 Db 1554 GAGACAGCAAGAAAGCCCTCTGGAACATGGAGAGATGTTGATGACCATTTGCCCATG 1613
 Qy 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysArgGln 660
 Db 1614 GTGGGCTCCAAGACCGTGTCCAGTGAAGATTTACTTCAACTCAAGAAAGAGCGAG 1673
 Qy 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680
 Db 1674 AACCTGACGAAATCTTTCAGACAGACAGCTAAAGATGAGAGAGAGAGAGAGAG 1733
 Qy 681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 700
 Db 1734 AGAAGAGAAAGAAAGACCCCAAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1793
 Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
 Db 1794 GAGGAGCAAGATGGAAGCATCAGCGCAAGTGCATAGAGAGAGAGAGAGAGAGAGAG 1853
 Qy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArg---GlyGluCysSerGlyPro 739
 Db 1854 GCAGAGCTTCACAGAGCTCTGGGAATGAGATTCCTCCAGATTTGGGAGTGCAGTGGCCA 1913
 Qy 740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla 759
 Db 1914 GCTGCTGCACAAACAGCTCTGATGAGAGTGTCCATCTCCCGCTTGAGAGAGAGAG 1973
 Qy 760 LysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAspGly 777
 Db 1974 AAGGAGACT-----GGGCTTAAACCACTGGGACCTGAGAGATTTGCCGCTGCCACC 2024
 Qy 778 ProProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThr 797
 Db 2025 CAAGCACTGTTCT-----CTTCCAGAGAAACCGGCACTAGCCCTGTGTGAGCTTCC 2078
 Qy 798 ProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAla 817
 Db 2079 CAGTCCCTGATCCAGTGGCCCAACATCCCAAGAGCTTCC---CATACCTTGGCGCA 2134
 Qy 818 ProProProValAlaProLysGluGluLysGluGluGluThrAlaAlaAlaProProVal 837
 Db 2135 CCCCAGGCTATCTGTGACAAGATGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2194
 Qy 838 GluGluGlyGluGluGluLysProProAlaAlaGluGluLeuAlaValAspThrGlyLys 857
 Db 2195 GAGGATGCCAAGAGAGAGAAATCTGAGCGAGAGAG-----ATCGATGTGGGAGAAA 2245
 Qy 858 Ala-GluGluPro-----ValLysSerGluCysThrGly 868
 Db 2246 GCCAG 2305
 Qy 868 uGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThrAla 887
 Db 2306 GAGAGCCGAG 2365
 Qy 887 ArgGluGlyAlaLeuLysAlaGluLysLysGluGlyLysSerGlyArgAlaThrThrAla 907
 Db 2366 TGAAGCAACCTTAAGTGAAG---AAGGCTGTAGCAAGAGAGAGAGAGAGAGAGAG 2422

Qy	2350	lMergllyserThrAspGlnTTPGluGluSerProProleuSerAlaAnaIapheAsPr	2370
Db	6617	CATGGGTAAATATGATGTCAGTGGAAAGACCCCCCGCGTGGCGCAATGCTTTAAACC	6676
Qy	2370	OLEuSnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyAR	2389
Db	6677	TCTGAATCCACGGCGCACTCTGCCGCTGTCTTATGCCATACCACTGCTGACGACG	6736
Qy	2389	gSerAspHisThrLeuThrSerProGlyGlyGlyGlyAlaIysAlaIysValSerGlyYArgPr	2409
Db	6737	GAGTACACACCACTCACTCGCCAGGTGAGGTGGAAAGCAAGGTCTCTGGACGAC	6796
Qy	2409	oSerSerArgYsAlaIysSerProAlaProGlyLeuAlaSerGlyAspArgProProse	2429
Db	6797	TAGACAGCGGAAAGCAAGTGGCCAGCACAGGCTTACGTCCTGGAGACGACCCCCCTTC	6856
Qy	2429	rValSerSerValHisSerGluGlyAspCysAnaArgTrgThrProLeuThrAsnArgVa	2449
Db	6857	TGTCCTCCTCACTACACTAGAGGGGGACTGCATGCGCCGAACCACTACCAACCGTGT	6916
Qy	2449	lTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProleuIleMe	2468
Db	6917	GTGGAGGACCGGCGCCCTCATGTGAGGGGTCCACGCGCATTTCCCTCAACCCCTTGATTAT	6976
Qy	2469	tArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerG	2488
Db	6977	GAGGTACAGGCAAGGTGTTCATGTGCTCCCGCCACCTGGCGCTTGGGAGGACGCGG	7036
Qy	2489	yProleuAlaGlyProHisAlaTrpAspGluGluProIysPProleuLeuCysSerG	2509
Db	7037	GCCCTTACCTGTGCTCCCAACGACGCTGGGATGAGAGCCCAAGCACTGTGTGTTCACA	7096
Qy	2509	nTyrGluThrLeuSerAspSerGlu	2517
Db	7097	GTATGAGACACTCTCGACAGCGAG	7121
RESULT 7			
LOCUS	CQ722208	6339 bp	DNA linear PAT 03-FEB-2004
DEFINITION	Sequence 8142 from Patent WO02068579.		
ACCESSION	CQ722208		
VERSION	CQ722208.1	GI:42283065	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Alignment Scores:			
Score:	2,39e-189	Length:	6339
Percent Similarity:	9776.50	Matches:	1958
Best Local Similarity:	75.36%	Conservative:	6
Query Match:	75.13%	Mismatches:	59
DB:	73.98%	Indels:	583
	6	Gaps:	15
US-09-522-753-5 (1-2517) x CQ722208 (1-6339)			
Qy	1	MeSerGlySerThrGlnLeuValAlaGlnThrTPdArgAlaThrGluProArgTyrPro	20

Db	1	ATGTGGGATTCACACAGCGCTGTGGCACACAGCGTGAAGGGGCACTGAGCCCCGCTACCG	60
QY	21	ProHisSerLeuSerTyrProValGlnIleAlaArgThriSthrAspValGlyLeuLeu	40
Db	61	CCCCACAGCGCTTTCCTACCCAGTCCAGATCGCCCCGAGCGCACACGAGCGTGGGCTCTG	120
QY	41	GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60
Db	121	GAGTACCAAGCACCACTCCCGACTATGACTCCCACTGTGTGCCCCGGCTTCATATCCAG	180
QY	61	ProGlnArgArgArgProSerLeuLeuSerGlnPheGlnProGlyAangIuArgSerGln	80
Db	181	CCCCACGCGCGAGAGGCCCTCCCTCTGTCTGTAGTTCACGCCCGGAATGAACGGTCCAG	240
QY	81	GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGluLeuGlyLeuSerGluMet	100
Db	241	GAGCTCCACCTCGCGGCAGAGTCCCACTCATCTCCGCCGAGCTGGGGAAAGTCAGAAATG	300
QY	101	GluPheIleGluSerTyrSArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro	120
Db	301	GAGTTCATTGTAACCAACAGCGCCCTCGCTAGAGTGTGCTGACCCCTGCTCGACCG	360
QY	121	SerProLeuLeuAlaThrGlnIleProAlaGlySerGlnAspLeuThrIlyAspArgSer	140
Db	361	TCAACCCCTCGCTGGCCACGGGCGCACCTCTCGGGATCTAAGACCTCCACCAAGACCGTAGC	420
QY	141	LeuThrGlyIlyLeuGlnProValSerProProSerProProHisSthrAspProGluLeu	160
Db	421	CTGACGGGCAAGCTGGAAACCGGTCTCTCCGCCCGCCGCACTGACCTCTGACCTG	480
QY	161	GluLeuValProProArgLeuSerTyrGluGluLeuIleGlnAsnMetAspArgValAsp	180
Db	481	GAGCTGTGTGCGGCACAGCTGTCCAGAGAGAGCTGATCCAGAACATGAGCGGTGGAC	540
QY	181	ArgGluIleThrMetValGluGlnGlnIleSerIlySerLeuIlySlySlyGlnGlnLeu	200
Db	541	CGAAGATCACCAATGATGAGACACAGATCTCTAAGCTGAAGAAAGAAAGACACACTG	600
QY	201	GluGlnGluAlaAlaIlyProProGlnProGluIlyAspProValSerProProIleGlu	220
Db	601	GAGGAGGAGGCTGCCAAGCGCCCGAGCGTGAAGCCCGTGTACCGCGCCCATGAG	660
QY	221	SerTyrSHiSArgSerLeuValGlnIleIleTyrAspGluAsnArgIlySlyAlaGluAla	240
Db	661	TCGAGACACCGCACCCCTGTGTCAATATCATCAACAGAAACGGAAGGCTGAAGCT	720
QY	241	AlaHisArgIleLeuGlnGlyLeuGlyProGlnAlaGluLeuProLeuTyrAsnGlnPro	260
Db	721	GCACATCGAATTCGTGAGAGGCTGTGGGCCCCAGGTGAGCTGCGCTGTACACACACCC	780
QY	261	SerAspThrArgGlnTyrHisGluAsnIleIlySileAsnGlnAlaMetArgIlySlyLeu	280
Db	781	TCCGACACCGCGGAGTATCATGAGAAATCAATAAATTAACGAGCGGATGCCGAAGAAGCTA	840
QY	281	IleLeuTyrPheIlySArgArgAsnHisAlaArgIlyGlnTrrPlySglnIlySphCyGln	300
Db	841	ATCTTGTAACCTCAAGAGAGGAATCAAGCTCGGAACAATGGGAGCAGAAATTCTGCAG	900
QY	301	ArgTyrAspGlnLeuMetGluAlaLeuGlnIlySlyValGluArgIleGluAsnAsnPro	320
Db	901	CGCTATGACCGACTCATGAGAGGCTGTGGAGAGAAAGGTGAGGCCATCGAGAACACCC	960
QY	321	ArgArgArgAlaIlySglnSerIlyValArgIlyTyrGlyIlySglnPheProGluIle	340
Db	961	CGGCGCGCGGCCCAAGAGAGCAAGTCCGCGAGTACTACAGAAAGAGTTCCCGAAGTC	1020
QY	341	ArgIlySglnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly	360
Db	1021	CGCAAGCAGCGCGAGCTGACAGAGACGCAATGCAAG--AGGGTGGCCAGCGGGGCAGTGG	1077
QY	361	LeuSerMetSerAlaAlaArgSerGlnHisGluValSerGluIleIleAspGlyLeuSer	380
Db	1078	CTGTTCATGTGTGGCGCGCCGAGGAGACGAGGTGTCAAGATCATGATGCTCTCA	1137

QY	381	GIUGINGIUAenleuGIULySGImeCargIuLeuAlaValIleProPromeIeUtyr	400
Db	1138	GAGCAGGAGAACTCGTAGAGAGAGATGGCCGAGCTGGCGGTGATCCGCGCATGCTGAC	1197
QY	401	AApAlaAaSGInGIArgIlelysheIleaaenMeCAnGIyLeuMetAlaAspProme	420
Db	1198	GACGTGACACGACAGCGATCAAGTTCATCAATGAAAGGAGCTTATGGCCGACCCATTG	1255
QY	421	LySValIYrLySaSPArgGIuValMeAenMeCTPSeGIuGIuGIuLySGIuThrPhe	440
Db	1258	AAGGTGTACAAAGACCGCCAGAGTCATGAACATGTGAGTGAAGCAGAGAGAGACTTC	1317
QY	441	ArgGIuLySPheMeGIuhiSProLySaenPheGIyLeuIleAlSeSPheleuGIuArg	460
Db	1318	CGGGAGAAgTTCATCAGCATCCCAAGAACTTTGGCTGATCGCATCTATCTCTGGAGAG	1377
QY	461	LySThrValAlaGIuCySeValleuTyrrTyrrIleuThrLyblySaenGIuAsnTyrrys	480
Db	1378	AAg---GTGAGTCGCTGC-----	1392
QY	481	SeIreuValArgArgSeTyrrArgArgIySGIySeGIuGIuGIuGIuGIuGIuGIu	500
Db	1392	-----	1392
QY	501	GIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu	520
Db	1392	-----	1392
QY	521	AePGIuLySGIuLySGIuLySGIuLySGIuLySGIuLySGIuLySGIuLySGIuLy	540
Db	1393	-----CCGCCCCATACCCCTTCGCTCCACC	1419
QY	541	AspLySGIuAePLeuLeuLyS-----GIuLySThrAaSPThrSeGIyLyaSPaSn	558
Db	1420	TCCGGGGCACTGGGTGAGACTCAAGGAGAAAGACAGACGACACTCGGGAGAGAACAC	1479
QY	559	AePGIuLySGIuLyAlaValAlaSeTyrrGIyArgLySThrAlaenSeGIuLyArgArg	578
Db	1480	GACGGAAGAGGGCTGTGGCTCCCAAGCCGCAAACTGCCACACGACGAGGAAGAGCC	1539
QY	579	LySGIyArgIleThrArgSeMeAlaAenGIuLyAaenSeGIuLyAlaIleThrPro	598
Db	1540	AAAGGCCCATCACCCGCTCAATGGCTTAATGAGGCCAACAGAGAGAGCCATCACCCC	1599
QY	599	GIuGIuSeIraGIuLyLeuAlaSeTyrrGIuLyAaenGIuLySeTyrrTrpGIuGIu	618
Db	1600	CAGCAGACCGCCGAGCTGGCTCCATGAGCTGAATGAGAGTTCGCTGACAGAAAGAA	1659
QY	619	GIuMeTGIuThrAlaLyblySGIyLeuLeuGIuhiSGIyArgAnTPSeAlaIleAla	638
Db	1660	GAATATGAAGAACGCCAGAAAGGTCTCTTGAAACAGCGCCCAACTGGTCGGCATCGCC	1719
QY	639	ArgMeTValGIySeTyrrThrValSeGIuCyblySaenPheTyrrPheAsnTyrrLyblys	658
Db	1720	CGGATGTGGGCTCCAGACATGTGTGGAGTGTAAAGACTTCTACTTCACTACAGAG	1779
QY	659	ArgGIuAenLeuAaSPGIuIleLeuGIuGIuhiSGIyLeuLybMeTGIuLySGIuArgAsn	678
Db	1780	AGGCAGAACTCGATGAGATCTTGACAGCAGCAAGCTGAAGATGAGAAAGAGAGGAC	1839
QY	679	AlaArgArgLyblySGIyLyblyAlaProAlaAlaAlaSeGIuGIuAlaIlePheProPro	698
Db	1840	GCGCGAGAGAAAGAAAGAAAGAGCCCGCGCGCGGCAACGAGGAGGCTGCACTTCCGCC	1899
QY	699	ValValGIuAePGIuGIuMeTGIuAlaSeGIyValSeGIyAaenGIuGIuMeTVal	718
Db	1900	GTGTGTGAGAGATGAGAGATGAGAGCGTCCGGCTGAGCCGAATGAGAGAGATGGTG	1959
QY	719	GIuGIuAlaGIuLyAlaLeuhiAlaSeGIyAaenGIuValProArgGIyGIuCySeGIy	738
Db	1960	GAGGAGGCTGAAGCCTTACATGCTCTGGGGAATGAG-----	1995

QY	739	ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla	758
Db	1995	-----	1995
QY	759	AlaLysAspThrGlyGlnAsnGlyProLysBProProAlaThrLeuGlyAlaAspGlyPro	778
Db	1995	-----	1995
QY	779	ProProGlyProProThrProProArgThrSerArgAlaProIleGluProThrPro	798
Db	1995	-----	1995
QY	799	AlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaPro	818
Db	1995	-----	1995
QY	819	ProProValaValaProLysGluGlnLysGluGlnGluThrAlaAlaAlaProProValaGlu	838
Db	1995	-----	1995
QY	839	GluGluGluGluGlnLysProProAlaAlaGluGluLeuAlaValaAspThrGlyLysAla	858
Db	1995	-----	1995
QY	859	GluGluProValaLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLys	878
Db	1995	-----	1995
QY	879	AspAlaGluAlaAlaGluAlaThrAlaGluGluGlyAlaLeuLysAlaGluLysGluGly	898
Db	1995	-----	1995
QY	899	GlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSer	918
Db	1996	-----GACAGGAGCTCCAGT	2010
QY	919	AlaThrCysSerAlaAspGluValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeu	938
Db	2011	GCCACCTGTCAGTGCAGACGAGGTGGATGGAGGCCGAGGGCGCGACAGAACCCTGCTCTG	2070
QY	939	SerProArgProSerSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln	958
Db	2071	TCCCCAAAGGCCACCTCTCCACCCCGACCTGGCGACCCCGGGCCAAATGCTTCACCCCCAG	2130
QY	959	LysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnVal	978
Db	2131	AAGCCACTGGACCTGAAAGCACTGAAAGCGAGCGAGCTGCATCTCCCCCATC-----	2184
QY	979	ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAla	998
Db	2184	-----	2184
QY	999	ProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer	1018
Db	2184	-----	2184
QY	1019	ProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGlu	1038
Db	2185	-----GACGCTGACCCCTCTCTGGGCCCA-----	2211
QY	1039	AlaGlnLysLeuProGlyAspProProCysAlaThrThrSerGlyLeuProPheProValaPro	1058
Db	2212	-----TGGACA-----	2217
QY	1059	ProArgGluValIleLysAlaSerProHisAlaAlaProAspProSerAlaPheSerTyrAla	1078
Db	2218	-----GACATGTGC	2226
QY	1079	ProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValaLeuProArg	1098
Db	2227	TTCACAGGTACCCCACTGCCCCCTGGAGGCTCAATGACACTGCCCGGCCGCTGCGCGC	2286
QY	1099	ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeu	1118

DB 2287 CCACCCACCATCTCCCAACCCGCTCCCTCATCTCCCTGCGCAAGCACCCGCGTCTTC 2346
QY 1119 GIATGAGLIIIEGLYALIIIESERGLNGIYMETSERVALGIDNEUHSIVALPROTYRSE 1138
DB 2347 GAGAGCGAAATAGTGCATCTCCCA----- 2373
QY 1139 GIUHSIALALYALAPROVALGIPROVALTHMETGLYLEUPROLEUPROMETAEPPRO 1158
DB 2373 ----- 2373
QY 1159 LYSLYSLEUALAPROPHESERGLYVALLYSGINGINGIUNLEUSERPROARGIYGINALIA 1178
DB 2373 ----- 2373
QY 1179 GLYPROPROGLUSERLEUGLYVALPROTHRAGINGIUALASERVALLEUARGLYTHR 1198
DB 2373 ----- 2373
QY 1199 ALALEUGLYSERVALPROGLYGSERILETHRLYSGIYLEPROSERTHRARGVALPRO 1218
DB 2373 ----- 2373
QY 1219 SERAPSERVALIETHRTYRARGIYSERILETHRHISGLYTHRPROALASPVALLEU 1238
DB 2374 -----GGCACGCGCAGCTGACGTCTTG 2394
QY 1239 TYLYSGIYTHRILETHRARGLILEIIEGLYGLUASPSEPROSERARGLEUAAPARGIY 1258
DB 2395 TACAGAGGACCATCTCCAGAGATCATCGCGAGAGACAGCCCGAGTGGCTTGACCGCGGC 2454
QY 1259 ARGGLUASPSELEUPROLYSGIYHISVALIIEITYRGLUGLYLYBALYSGIYHISVALLEU 1278
DB 2455 CGGAGAGACAGCGCTGCCCAAGGGCCACGTTCATCTACAGAGGCAAGAGGCGCACGTCTTG 2514
QY 1279 SETTYRGLUGLYGLYMETSERVALTHRGINYSERLYSGIUALSPGLYARGSERSE 1298
DB 2515 TCCTATGAGGGTGGCATGTCTGTGACCCAGTGTCTCAAGAGAGACGCGAAGACCACTCA 2574
QY 1299 GLYPROPROHISGLIUTHRALAALAPROLYSARGTHRTYRASPMEWETGLIYASVAL 1318
DB 2575 GAGACCCCGCATAGACGCGCGCGCCCGCCAGCGCACTTAGCATGATGAGGCGCGCGTG 2634
QY 1319 GIYARGALIIIESERSERALASERILEGLIYLEUMETGLYARGALIIIEPROPROGLI 1338
DB 2635 GCGAGAGCCATCTCTCAGCGACGATCGAAGT----- 2667
QY 1339 ARGHISERPROHISHIILEUUYSGIUGINHISIEILEARGIYSERILETHRGINGLY 1358
DB 2667 ----- 2667
QY 1359 ILEPROARGSETYRVALGIUALAGINGIUALSPTYRLEUARGVALIYLSLEULEU 1378
DB 2668 -----GATGACGAGGAGAGACTTACTCTTC 2694
QY 1379 LYSARGIUGLYTHRPROPROPROPROPROSERARGASPLEUTHRGUALATYLYS 1398
DB 2695 -----GATGCGCC-----TGCGGGCGGCTTCAAG 2721
QY 1399 THRGINALALEUGLYPROLEUYLSLEUYSPROALHISGLIYLYLEUVALIATHRVAL 1418
DB 2722 AGCGAGGCGCTGGCGCCCTGAGCTGAGCCGCGCATGAGGGCGCTGGGGCGCAAGG 2781
QY 1419 LYSGLUALAGIYARGSERILEHISGLIUILEPROARGIUGIULEUARGHISHTHPROGLI 1438
DB 2782 AAGGAGGCGGCGCTTCATCATGAGATCCCGCGAGAGCTTGGCGCACAGCGCGAG 2841
QY 1439 LEUPROLEUALAPROARGPROLEUYSGIUGLYSERILETHRGINGLYTHRPROLEUY 1458
DB 2842 CTGCGCCCTGGCGCGCGCGCTCAAGAGAGGCTTCATCACGAGGAGCACCGCGCTCAAG 2901
QY 1459 TYRASPETHGIALASERTHRTYRGLYSETYRLEUYSHISAPRYALARGSERLEUIIEGLY 1478
|||

DB 2902 TACGACACCGGCGGCTCCACCTGCTCAAAAAGCACAGACTACGCTCTCATCGCG 2961
QY 1479 SERPROGLYARGTHRPHERPROVALHISPROLEUAPRYALMETALASPALARGALA 1498
DB 2962 AGCCCGGCGGAGCGTTCCACCGGTGACCGCGTGGATGTATGCGGAGCGCGGCGCA 3021
QY 1499 LEUGIUALGALACYETTYRGLIUGIUSERLEUYSPSERARGPROGLYTHRALASERSE 1518
DB 3022 CTGGAACGTGCTGTACGAGAGAGGCTGAAAGACCGCGGAGCGCGAGCGCTCG 3081
QY 1519 GLIYLSERILEALARGGLYVALAPROVALIIEVALPROGLIUEUGLYLYSPROARGIN 1538
DB 3082 GGGGGCTCATTTGCGCGCGCGCGCGCGCTTGCTGACTGGGCAAGCGCGCGAG 3141
QY 1539 SERPROLEUTHRTYRGLUASPRIISGLYALAPROPEALAGIYHISLEUPROARGIYSE 1558
DB 3142 AGCCCGCTTACCTTAGAGAGACACGCGGGGACCTTTGCGGCGCACTCCACGAGGTTG 3201
QY 1559 PROVALTHMETARGIUPROTHRPROARGLEUGINGIUGLYSERILEUSERSE 1578
DB 3202 CCGTGACCAACGCGGAGGCCACGCGCGCTGACAGAGGCGAGCGCTTCGTCAAGCAAG 3261
QY 1579 ALASERGINASPARGLYLYLEUTHRSETHRPROARGIUIIEALALYSESPROHISSE 1598
DB 3262 GCATCCAGAGACGAAAGCTGACGTGACGCTCTGTAGATGCGCAAGTCCCGCACAGC 3321
QY 1599 THRVALPROGLIUIHISBPROHISBPROHISSEPROTYRGLIUIHISLEUARGIYVAL 1618
DB 3322 ACCGTGCCGAGACACACACACACCCCATCTGCGCTTAGAGACACTGCTTGGGCGGTG 3381
QY 1619 SERGLYVALASPLEUTHRTYRSGERHISILEPROLEUALAPHASPPROTHRSE 1638
DB 3382 AGTGCGTGAGACTGTATGCGAGCACATCCCCGCGCTTGACCCCACTTCATATCC 3441
QY 1639 ARGGLYILEPROLEUASPALIAALIAIATYRGLYLEUPROARGHISLEUALAPROAN 1658
DB 3442 CGCGGATCTCTTTGGACGACCGCGCTGCTACTACCTGCGGACACTGCGCCCAAC 3501
QY 1659 PROTHRTYRPROHISLEUUTHYRPROPTYRILEUIIEARGIYTYRPROAPRTHRALIA 1678
DB 3502 CCCACTACCGCACCTGTATCCACCTTACCTGCGGCTTACCGGCGTACCGGCGGCG 3561
QY 1679 LEUGIUALSPARGIINTHRIIEIIEHISAPRYTILETHSERGLIIMETHISHSAP 1698
DB 3562 CTGAGAAACCGGACGACATCATGATCATCATCTCCAGAGATGACCAAC 3621
QY 1699 THRALTTHRALMETALAGIINARGALIASPMETLEUARGIYLEUSERPROARGIUSER 1718
DB 3622 GCGGCGACCGCATAGGCCAGGAGCTGATAGTGAAGGCGCTCTCGCCCGGAGTGC 3681
QY 1719 SERLEUALALEUASNTYRILAAIAGLYPROARGIYILEIASPLEUSERGINVALPRO 1738
DB 3682 TCGGTGGACATACATAGCGTCCGGGTCCCGAGGATCATGACCTGTCCCAAGGCA 3741
QY 1739 HISLEUPROVALLEUVALPROPROTHRPROGLYTHRPROALATHRALMETASPARGLEU 1758
DB 3742 CACTGCGCTGTCTGTGCGCCCGGACACAGGACCGCCAGCGCATGAGCGCTT 3801
QY 1759 ALATYRLEUPROTHRALAPROGLINPROPHESERSEARGHISESERSEPROLEUSER 1778
DB 3802 GCTTACTTCCACCGCGCGCGCCAGCGCTTACAGACCGCGCACAGACTCCCACTCTCC 3861
QY 1779 PROGLIYLYPROTHRHISLEUTHRLYSPROTHRTHRTHRSESERSEGLIARGIUALG 1798
DB 3862 CAGAGAGTCAACACACTTGACAAACCAACACACACAGCTCTCGHCGAGGCGGAGCA 3921
QY 1799 ASPARGASPARGLIUALARGASPARARGIUALARGIULYSE 1818
DB 3922 GACCGGAGTCAAGAGGCGGATCGGAGCGGAGAAATGATCTTCAGCTCAC 3981
QY 1819 THRTHRVALGIUHSIALAPROILETRPARGPARGIYTHRGINGIUSERSE 1838
DB 3982 ACAGCGTGAACGACGACCATCTGAGACTGTGAGGCACTGAGGCAACGCCCCCA 4041
|||

Qy 1839 GlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHis 1858
Db 4442 GCTCCGGAGACTCTTGTGGCCGCAAGAGGCTCCCTGCTGATGCACTGACTGTAC 4101
Qy 1859 GlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu 1878
Db 4102 CAGACTCGCCACTCTCCCTCGGACCCAGATGCCCTCCAGCAGAGACCAAGTGTCTT 4161
Qy 1879 HisAsnThrGlyMetLeuGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
Db 4162 CACAACAGGCAAGAGGATATCATCACCGCTGGAGGCCAGACGACCGCTCTG 4221
Qy 1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
Db 4222 AGGTCCACTCTCACTCTCTCACTCCGTTGCGCCGCTGCACATTCACCTGCCACACC 4281
Qy 1919 CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
Db 4282 TGCCCACTGGGCGGACCTCTGATGGGGTCTACCTCACTGAGAGCCGCTCTGCTG 4341
Qy 1939 ProLeuGlnAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
Db 4342 CCCAAGAGGCCCCCGGGTCCCGCGCAGAGGCGCCGAGCAGACACCGGCAATGCC 4401
Qy 1959 PheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGly 1978
Db 4402 TTCTTCCCAAGCCCCCAAGCCGCTCGGGCTGAGGCCGCTCTCTCCCAAGAGGGC 4461
Qy 1979 SerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPro 1998
Db 4462 TCGGAGCCCGGCCCCCTAGTCTCTCTGTCCTGCGCACCGCACATGCGCCGACCCCT 4511
Qy 1999 AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAla 2018
Db 4522 GCGAAGAACCTCGCACTCAACAGCAGCCGCGAGCCGCGCGCCACCTGCTCGGCC 4581
Qy 2019 SerAspProHisArgGlyLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeu 2038
Db 4582 TCGGACCCGCGCGGAAAGACTCAAAAGTAAACCTTTTCATCCAGAACTGAACTC 4611
Qy 2039 ArgSerLeu----- 2041
Db 4642 CGTCTCTGGGTAGAACCACTGACCGGCGCACTTCATAGACGATATCATCGT 4701
Qy 2041 ----- 2041
Db 4702 CAAATTGCTCAGATAAGGGCGCGAAGAGAGTGCCTGGCCAAAGGCTCCCTCGC 4761
Qy 2042 ---GlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
Db 4762 GATGGTATCAACGGCAGACGACTACGCCCGGAGGGGTGAGCCGCTGACGCTGTAGC 4821
Qy 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer 2080
Db 4822 TCACCCAGTCTGACCGACGACAGAGGGCTCCCAAGCACTCGAAGAGCTCGACAAAGC 4881
Qy 2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyValAla 2100
Db 4882 CACTGAGGGGAGGCTGCGGCCCAAGCAGCAGCCCGGTGAAGCTTGGCGGGAGGCC 4941
Qy 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
Db 4942 GCCCACTCCCAACACTCGGCGGCTGCTGAGAGCCAGCCCTGTCTCAAGCCGCTCTC 5001
Qy 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValAlaThrLeuAlaGlnHisIleSer 2140
Db 5002 CAGACCGCCCGAGGGGTCAAAAGTCAACAGCGGGTGTCACTCGGCCAGCATCATAGT 5061
Qy 2140 ----- 2140
Db 5062 GTAACTACGGCTTCTGTGCTGCTGTGTCACTTTGCACTGGGGGACCAAGCCTGGA 5121

Qy 2140 ----- 2140
Db 5122 GAGGGATGGGGAACCCACAGGCCCTTCTGTCTGCGGGGTGGCTGGGGATCCAGGGC 5181
Qy 2141 -----GluValIleThrGlnAspTyrThrArgHisHis 2151
Db 5182 ATGGCGTGGGGGATCCAGAGCGGTGGAGGTATCAACAGGACTACACCGCCAGCACAC 5241
Qy 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCys 2171
Db 5242 CCACAGAGCTCAGCGACCCCTGCGCCGCCCTCTACTCTCTCTCTGGGGCCAGCTGC 5301
Qy 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGly 2191
Db 5302 CCGTCTTGACCTCGCGCCGCCCAACGAGCTTCACTTCCGCCGCCGAGCAATGGT 5361
Qy 2192 AlaProAlaArgLysProHisSerGluGlyLysArgSerProGluProAsnLys 2211
Db 5362 GCCCGGCGCGTGGCTCCCCCAAGAGGGGGCAAGAGTCTCCAGAGCCAAACAG 5421
Qy 2212 ThrSerValLeuGlyGlyGlyLysArgLysIleGluProValSerProProGluGlyMet 2231
Db 5422 ACGTGGTCTTGGTGGTGGTGAAGAGGTATTAACCTGTCTCCACCGAGGGCATG 5481
Qy 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
Db 5482 ACGGAGCGACGCACTCCCGAGTCTGTGTACCCGCTGTACCGGATGGGGAAACAG 5541
Qy 2252 ThrGluProSerArgMetGlySerLysSerProGlyValAsnThrSerGlnProProAlaPhe 2271
Db 5542 ACGGAGGCC---AGATGGGCTCCAAGTCTCAGGCAACACAGCCAGCCGACCTTC 5581
Qy 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnIleAsn 2291
Db 5599 TTCAGCAAGCTGACCGAGAACCTCCGCAATGATGATCAAGTCAAGAAAGAGATCAAC 5658
Qy 2292 LysLeuLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
Db 5659 AAGAGCTGAACACCAACACCGGAATGAGCTCGAATACATATTCAGCCAGCTGGGACG 5718
Qy 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
Db 5719 GAGATCTTCAATATGCGCCGCTACACCGGAACAGGCTTATGACCTATAGAACCCAGGCG 5778
Qy 2332 ValGlnGlnHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMet 2351
Db 5779 GTCCAGAACATCCAGACCAACATAGGGCTGAGGCCATTAATTTGAAGGCACTCATG 5838
Qy 2352 GlyLysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
Db 5839 GTTAATATGACCAAGTGGGAAGAGTCCCGCGCCTCAGCCCAATGCTTTAACCTCTG 5898
Qy 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 5899 AATGCCAGTCCAGCTGCGCCGCTGTATGCCATTAACCGCTGTGACGAGCAGAGTAC 5958
Qy 2392 HisThrLeuThrSerProGlyGlyGlyLysValLysValSerGlyArgProSerSer 2411
Db 5959 CACACACTCACTCGCCAGGTGGCGGGAGGAGGCAAGGTCTCTGCAACACCCAGCAGC 6018
Qy 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 6019 CGAAAGCCAAAGTCCCGCGCCCGGCGCTGTGATCTGGGGAACGGGCACTCTGTCTCC 6078
Qy 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGlu 2451
Db 6079 TCAGTCACTCGGAGGAGAGACTGCAACCCCGGACCCCTCAACCAACCGCGTGGGAG 6138
Qy 2452 AspArgProSerSerArgLysSerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db 6139 GACAGGCCCTCGTCCGAGGTTCACGCCCATTCCTTACAAACCCCTGTATCATAGCGGCTG 6198
Qy 2472 GlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491

[illegible]

exon		4483..4620 /gene="SMRT"	
ORIGIN		/note="alternatively spliced insert"	
Alignment Scores:			
Pred. No.:	3,33e-151	Length:	5989
Score:	7885.00	Matches:	1510
Percent Similarity:	94.34%	Conservative:	7
Best Local Similarity:	93.91%	Mismatches:	14
Query Match:	59.67%	Indels:	77
DB:	9	Gaps:	7
US-09-522-753-5 (1-2517) x HSU37146 (1-5989)			
QY	984	ProProAaTgGiuaaPaAlaAlaProThLySerProAlaProProAla-----	998
DB	168	CCACCAAGA-----TCACCAATAGACCAAGACCCCTCGATGGAATATCATAT	218
QY	999	ProProProPro-----	1002
DB	219	CCGAGACCAACCTGTGATCATAGCCATCTGCATTCTGATGGCTCCATGATCTATT	278
QY	1003	-----GlnaLeuGlnProGluSerAaPaAlaProGlnGlnProGly	1016
DB	279	GCAAGATCCCAAGTTTCATCAGCACTGAACCAACATCCCTGTGTAACGCTCCAGST	338
QY	1017	-----SerSerPro-----Arg	1020
DB	339	CTCCCTCTTCTGTATAGCCATGAGGCTCTCCCTCTAGAGTGATGATGAGAGAAGA	398
QY	1021	GlyLySerAlaGser-----ProAlaProProAla-----	1030
DB	399	GAAAGATTTCGAGCTCTGTACCAACCCGCGCTGTCTCGGAGAGGGGGAGTGGT	458
QY	1030	-----	1030
DB	459	CCTGACGAGGGCTCATATCATATCTCTTCTAATGAGAGGCATGGAGCGCCACCC	518
QY	1031	-AspLysGlnAlaPheAlaAlaGlnAlaGlnLysLeuProGlyAaProProCysTPrTh	1050
DB	519	CGAACAGGAGCGCTTCGACCGGAGCGCCAGAACCTCGTGGGAGACCCCTTCGTGAC	578
QY	1050	rSerGlyLeuProPheProValProProArgGlnValIleValAlaSerProHISAlaPr	1070
DB	579	TTTCGGGCTGCTTCCCGTCCCGCCCGGAGAGTGAAGCTTCCCGCATGCCCC	638
QY	1070	oAspProSerAlaPheSerTyrAlaProProGlyAlaProLeuProLeuGlyLeuHisAs	1090
DB	639	GGAACCCCTCAACCTTCTCTCAAGCTCCACCTGTCACCACTGCCCCGCGCTCCATGA	698
QY	1090	pThAlaArgProValLeuProArgProProThIleSerAsnProProProLeuIleSe	1110
DB	699	CATGCGCGCGCGCTTCGCGCGGCCCAACCATCTCCAAACCGCTCCCTCATCTC	758
QY	1110	rSerAlaLysHisProSerValLeuGlnArgGlnIleGlyAlaIleSerGlnGlyMetSe	1130
DB	759	CTTGTGCCAAGACCCCAAGCTCTCTCGAAGGCAATAGTGCCATCTCCCAAGAAATGC	818
QY	1130	rValGlnLeuHisValProTyrSerGlnHisAlaLysAlaProValGlyProValThMe	1150
DB	819	GGTCACAGCTCACCTCCGCTACTAGAGCAATGCCAAGGCCCGCGGTGGCGCTGTACCAT	878
QY	1150	tGlyLeuProLeuProMetAaProLysLysLeuAlaProPheSerGlyValLysGlnG	1170
DB	879	GAGGCTGCGCTTCGCCATGAGCCCAAAAGCTGGCACCTTTCAGCGAGTGAAGCAGGA	938
QY	1170	uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThAlaGl	1190

Db 939 GCAGCTTCCCGACGGGGCCAGGCTGGGCGCACCGAGAGCGCTGGGGTGCCACAGCCCA 998
Qy 1190 ngluaSerValIleuArgIyThraIaleuGlySerValProGIyGlySerIleThrIy 1210
Db 999 GGAAGCCTCCGTGCTGAGAGGAGACAGCTCGGGCTCAGTTCCGGGGCGAAGCATCACAA 1058
Qy 1210 agIyIleProSerThraRgValProSerAaspSerAlIleThrIyRargIySerIleTh 1230
Db 1059 AGGCATTCACAGCACCGGGGCGCTCGACAGCGCATCACATACCGCGGCTCATCAC 1118
Qy 1230 rHleGIyThProAlaaspValIleuTyLyGIyThIleThraRgIleIleGIyIuas 1250
Db 1119 CCACGGCAGCGCAGCTACGCTCGTCAAGGCGACCATCACAGAGTCAATCGCGAGGA 1178
Qy 1250 pSerProSerArgIleuAaspRargIyAArgGlyAaspSerIleuProIyagIyIleValIleTy 1270
Db 1179 CAGCCGAGTCGCTTGAGACGGGCGGGAGAGACAGCTCGCCCAAGGGCCACGTCATCTA 1238
Qy 1270 rGIyGIyLyIleagIyHleValIleuSerTyRGIyGIyGIyMeSerValThrgInCyse 1290
Db 1239 CGAAGCGAABAAGGCCACGCTTGTCTCTAGAGGTCGATGCTGTGACCCAGTCTC 1298
Qy 1290 rLyGIyAaspGIyAArgSerSerGIyProProHleGIyThraIaAlaProIyAArgTh 1310
Db 1299 CAAAGAGAGCGAGAAAGAGCTCAGAGACCCCCCATGAGACGGCGCCCAAGCGCAC 1358
Qy 1310 rTyTAaPMeMeGIyGIyAArgValGIyAArgAlaIleSerSerAlaSerIleGIyGIyLe 1330
Db 1359 CATAGCATGATGAGAGGCGCGGCGAGAGAGCATTCCTCGACCGCATCGAAGGCTCT 1418
Qy 1330 uMeGIyAArgAlaIleProProGIyAArgHleSerProHleHleuIyagIyInHleH 1350
Db 1419 CATGGGCGGTGCATCCCGCGAGAGGACACAGCCCCACACTCAAAAGACAGCACCA 1478
Qy 1350 gIleagrgIySerIleThrgInGIyIleProAArgSerTyValGIyAlaGIyInIuApyTy 1370
Db 1479 CATCCCGGGTCCATACACAAAGGATCCCTCGTCTCTAGAGGACAGAGAGCAT 1538
Qy 1370 rIleuAArgAArgIyAlaIleuIleuIyAArgGIyGIyThProProProProProse 1390
Db 1539 CTTGGCTCGGAGGCCAAAGCTCTTAAAGCGGAGGAGCGCTCCGCGCCACCGGCTTC 1598
Qy 1390 rArgAaPLeuThGIyAlaIleTyLyThThrgInAlaIleuGIyProIeIyIleuIyAProAl 1410
Db 1599 AGCGGACCTGACCGAAGGCTTCAAGAGCGAGGCTCGGGCCCTGTAAAGTGAAGCCGCG 1558
Qy 1410 aHleGIyGIyLeuValAlaThraValIyGIyAlaGIyAArgSerIleHleGIyIleProAr 1430
Db 1659 CCATGAGGGGCTGTGGCCACGGTGAAAGAGCGGGCGGCTTCATCATGATCCCGCG 1718
Qy 1430 gGIyGIyIleuAArgHleThProGIyIleuProIeIuAlaProAArgProIeIyagIyIse 1450
Db 1719 CGAGGAGCTGGGACACAGCGCCAGCTGCGCTCGGCGCCGCGCTCAAGAGGGGCTC 1778
Qy 1450 rIleThrgInGIyThProIeIySerTyRArgThrgIyAlaSerThThGIySerIyIy 1470
Db 1779 CATCAGCAGAGGACCCCGCTCAAGTACACACCGGCGCTCCACATGAGTCCAAA 1838
Qy 1470 aHleAaPValAArgSerIleuIleGIySerProGIyAArgThPhaProProValHleAProIe 1490
Db 1839 GCACGACGTACGCTCCTCATCGGAGCCCGGCGGAGCGTCCCAACCGTCGACCGGCT 1898
Qy 1490 uAaPValMeAlaAaPAlaArgAlaIleuGIyAArgAlaCyTyGIyGIyIleuIyIyIse 1510
Db 1899 GGAATGATAGCGCGAGCGCCGGGCACTGGAACGTGCTGTACAGAGAGAGCTGAGAG 1958
Qy 1510 rArgProGIyThraIaSerSerSerGIySerIleAlaAArgGIyAlaProValIleVa 1530
Db 1959 CCGGCGACGGGACCGCAGACAGCTCGGGGGGCTCATTTGGCGCGGCGCCCGGCTCATTTGT 2018
Qy 1530 lProGIyIleuGIyLyAaPArgIyInSerProIeIuThTyRGIyAaPHeGIyAlaProPh 1550

Db 2019 GCCTGAGCTGGTAAAGCCGGGAGAGCCCTGACCTATGAGACCAAGGGGACCCCTT 2078
Qy 1550 eAlaGIyHleIleuProArgIyIleSerProValThMeAArgGIyProHraProArgIeGI 1570
Db 2079 TGCGGGCACCTCCACAGAGTTTCGCCGTGACCATCGGAGAGCCAGCGCGCTGCA 2138
Qy 1570 nglIySerIleuSerSerSerIyValaSerGIyAaPArgIyIleuThSerThProAr 1590
Db 2139 GAGAGGAGCCTTTGCTCGACAGAGCATCCAGAGCCGAAGGTGACGTGACGCTCG 2198
Qy 1590 gGIyIleAlaIySerProHleSerThraIProGIyHleHleAProHleAProIleSerPr 1610
Db 2199 TGAGATCGCAAGTCCCGCAGAGACCGCGCCGAGACACCCACCTCCTCGCC 2258
Qy 1610 cTyRGIyHleIleuAArgIyValaSerGIyValaAaPLeuTyAArgSerIleIleProIe 1630
Db 2259 CATAGACACTCTTCGGGGCGTGTAGGCGTGAACCTGTATCGGACCAATCCCCCT 2318
Qy 1630 uAlaPheAaPProThSerIleProAArgIyIleProIeIuAaPAlaAlaIaIaTyTy 1650
Db 2319 GGCCTTGACCCACCTCCATACCCCGGCGCATCCCTGTGAGCGACGCGCTGCTACTA 2378
Qy 1650 rIleuProAArgHleuAlaProAaPProThTyProHleIleuTyTyProProTyIleuI 1670
Db 2379 CTTGCCCGGACACTGTGGCCCCCAACCCCATCCGCACTGTATCCACCTAATCTCAT 2438
Qy 1670 eAArgGIyTyTyProAaPProThAlaIleuGIyAaPArgInThIleIleAaPProTyI 1690
Db 2439 CCGGGCTACCCCGACAGCGGGCGCTGAGAAACGAGACCATATCATATCATATCAT 2498
Qy 1690 eThrSerGIyInMeChleHleAaPThraIaThraIaMeValaGIyAArgIaAaPMeCle 1710
Db 2499 CACTCGACAGACATGACACACAAACAGCGCACCGGCATGCGCCAGCGAGCTATATGCT 2558
Qy 1710 uAArgIyIleuSerProAArgIyIleuSerIleuAlaIleuAaPArgIyAlaGIyProAArgI 1730
Db 2559 GAGGGGCTCTCCCGCGGAGTCTCGGTGACATCACTAGCGTGGGGTCCCGAGG 2618
Qy 1730 yIleIleAaPLeuSerGIyValProHleIleuProValIleuValProProHraProGIyTh 1750
Db 2619 CATCATGACCTGTCCCAAGTGCACACCTGCTGTGTGTCGTCGCCCCGACACAGCAC 2678
Qy 1750 rProAlaThraIaMeAaPArgIleuAlaTyIleuProThraIaProGIyAaPProHraSerSe 1770
Db 2679 CCAAGCAGACGATGAGACGCTTGCCTTACCTCCCAACGCGCCCAAGCCTTACAGCAG 2738
Qy 1770 rArgHleSerSerSerProIeIuSerProGIyGIyProThraHleuThIyAaPProThTh 1790
Db 2739 CCGCCACAGACGCTCCCACTCTCCCAAGAGGTCCAAACACTTGACAAAACCAACAC 2798
Qy 1790 rThrSerSerSerGIyAArgGIyAArgAaPArgAaPArgGIyAArgAaPArgAaPArgIyA 1810
Db 2799 CAGTCTCTGTCCGAGCGGAGAGAGACCGGATTCAGAGCGGAGACCGGAGTTCGAGAG 2858
Qy 1810 gGIyIySerIleIleuThSerThThraValGIyHleAlaProIleIleThraPProGI 1830
Db 2859 GGAAGATCATCTCTCATCTCACACAGCGGTGAAGACGACACCATGTGAAGACTGG 2918
Qy 1830 yThrgInGIyInSerSerGIySerSerGIySerGIyGIyGIyGIySerSerAr 1850
Db 2919 TACAGAGCAGACAGCGGAGAGCGGAGCGGCGGGGTGGGGGAGAGCAGCAGCGG 2978
Qy 1850 gProAlaSerHleSerHleAlaHleGIyHleIleuIleuIleuIleuIleuIleuIleuIleu 1870
Db 2979 CCGCGCTCCACTCCATGACCAACAGCATCGCCATCTCCCTCGGACCCAGGATGC 3038
Qy 1870 aleuGIyInIaArgProSerValIleuHleAaPThraIaMeCleGIyIleIleThraIaVa 1890
Db 3039 CTTCAAGAGAGAACCAAGTGTCTTCAACAACAGGCAAGAGGATATCATACGCTGT 3098
Qy 1890 lGIyProSerTyTyProThraIleuAArgSerThSerThSerSerProValAArgProAl 1910
Db 3099 GAGGCCAGAGCCACCGTCTGTGAGTCCACTTCCACTCTCAACCGGTTTCGCCACAG 3158

QY 1910 aAlaThrPheProProAlaThrHisCysProLeuGlyIleThrLeuAspGlyValIleTyrPr 1930
 Db 3159 TGGCAATTCCTCCACCTGGCACTCCCACTGGCGGCACTCTGAGTGGGGTCTAACCC 3218
 QY 1930 oThrLeuMetGluProValLeuLeuProLysGlyValAProArgValAlaArgProGluAr 1950
 Db 3219 TACCCCTCAAGAGACCGCTCTGCTGGCCCAAGAGAGCGCCCGGCTGCGCGCCAGAGGG 3278
 QY 1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGly 1970
 Db 3279 GCGCCGAGAGACACCGGCATGCTTCTCCGCAAGCCCCCAAGCCCGCTCGGGCTGGA 3338
 QY 1970 uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGly 1990
 Db 3339 GCGCCGCTCTCTCCCGCAAGAGGGCTCGAGCCCGCGCCCTAGTGTCTCTGTCTCTGG 3398
 QY 1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
 Db 3399 CCACGCCACCATGCGCCGACCCCTCGAAGAACTCGACCTCACCGCCAGCCCGGA 3458
 QY 2010 pProProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysPr 2030
 Db 3459 CCGCGCGGCGCACTGCTCGGCTCGGACCCGCAACCGGAAAGAACTCAAGTAAGTAAACC 3518
 QY 2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyIleYHisGlySerSerTyrSerPr 2050
 Db 3519 CTTTTCATCCAGAACTGGAACTCCGTTCTCTGGGTTACACGGAGAGAGCTACAGCCCC 3578
 QY 2050 oGluGlyValGluProValSerProValSerProSerProLeuThrHisAspLysGlyLe 2070
 Db 3579 CCAAGGGGTGGAGCCCGTGAAGCCCTGTGAGCTCACCAAGTCTGACCAAGAGGGGGCT 3638
 QY 2070 uProLysHisIleuGluLeuAspLysSerHisIleuGluGlyIleuLeuArgProLysGly 2090
 Db 3639 CCCCAAGCACTGGAGAGCTCGCAAGAGCCACTGAGAGGGAGACTGGGGCCCAAGGA 3698
 QY 2090 nProGlyProValLysLeuGlyIleValAlaHisIleuProHisIleuArgProLeuPr 2110
 Db 3699 GCCAGGCCCGCTGAAAGCTTGGCGGGAGGCGGCCCACTCCCACTGCGGCGCTGCGC 3758
 QY 2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGly 2130
 Db 3759 TAGAGGCCAGCCCTCGTCCAGCCGCTGCTCCAGACCGGCCCAAGGGGTCAAGAGTCAACA 3818
 QY 2130 nArgValValThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspThrThrArgHis 2150
 Db 3819 GCGGGGTGTACCTTGAGCCCAAGCAATCAGTGAAGTCAATCAACAGGACTACACCGGGA 3878
 QY 2150 sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe 2170
 Db 3879 CCAACCAAGCACTCAAGCGCACCCCTGCGCGCCCTCTTACTCTTCCCTGGGGCCAG 3938
 QY 2170 rCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHis 2190
 Db 3939 CTGCCCCGTCGTGAGACTCCGCGCCCAACCACTGACTCTTACCTCCGCGCCGAGACA 3998
 QY 2190 sGlyAlaProAlaArgGlySerProHisSerGlnGlyIleLysArgSerProGluProAs 2210
 Db 3999 TGGTCCCGCGGCGCGTGGCTCCCCCAAGCAAGGGGGGCAAGAGGCTTCAAGGCAACA 4058
 QY 2210 nLysThrSerValLeuGlyIleGlyIleuAspGlyIleGluProValSerProProGluGly 2230
 Db 4059 CAAGAGCTCGGTCTTGGTGTGTGTAGAGACGGTATTGAACCTGTGTCCCAACGAGAGG 4118
 QY 2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyIle 2250
 Db 4119 CATGACGAGAGCCAGGCACTCCCGAGAGTGTGTGTACCCGCTGTGTACCGGAGATGGGGA 4178
 QY 2250 uGlnThrGluProSerArgMetGlySerLysSerProGlyLysThrSerGlnProProAla 2270
 Db 4179 ACAGACGAGGCCAGAGATGGGCTTCAGAGTCTCCAGGCAACCAAGCAGCGCGCAGC 4238

QY 2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlyGluIle 2290
 Db 4239 CTTCTTACAGAACTGACCGAGAGCACTCCGCCATGTGTCAAGTCCAGAGAACAGAGAT 4298
 QY 2290 eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
 Db 4299 CAACAGAAAGCTGAACACCCACCAACCGGATAGCTGTGATATCAATATACGACAGCTGG 4358
 QY 2310 yThrGluIlePheAsnMetProAlaIleThrGlyLeuMetThrTyrArgSerGly 2330
 Db 4359 GACGAGAGATCTTCATATAGCCCGCATCACCGGAACAGGCTTATAGACTATAGAACCA 4418
 QY 2330 nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGlyValIleIleArgValAlaLe 2350
 Db 4419 GCGGCTGACAGAACTATGACGACCAACATGGGGCTGGAGGCCATATTAAGAAAGCACT 4478
 QY 2350 uMetGlyLysTyrArgProGlnThrGluGluSerProProLeuSerAlaAsnAlaPheAsnPr 2370
 Db 4479 CATGGGTAAATATGACCACTGGGAAAGATCCCGCGCTCAAGCCCAATGCTTTTAACC 4538
 QY 2370 oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe 2390
 Db 4539 TCTGAATGCCATGTCAGCTGCGCTGCGCTGTATGCTTATGCTGACGAGCGAGAG 4598
 QY 2390 nAspHisThrLeuThrSerProGlyGlyIleGlyLysAlaLysValSerGlyArgProSe 2410
 Db 4599 TGACCACACACTCACCTCGCCAGTGGCGCGGAGGCAAGGCTCTTGGAGAACCCAG 4658
 QY 2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
 Db 4659 CAGCGAAAGCCAAAGTCCCGGCGCCCGGCTGGCACTGGGAGACCGGCCACTCTGT 4718
 QY 2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
 Db 4719 CTCTCATGTCATCGAGGAGAGACTGCAACCGCGGAGCGGCTCAACCAACGCGGTGG 4778
 QY 2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr 2470
 Db 4779 GGAGGACAGGCCCTGCTCGGAGTTCACAGCATTCCTTCAACCCCTTATCATCTGG 4838
 QY 2470 gLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyPr 2490
 Db 4839 GCTGACGGCGGAGTGTATGCTTCCACCCCAACCGGAGCTTCCCGCGGAGACGGGGC 4898
 QY 2490 oLeuAlaGlyProHisHisAlaTyrAspGluGluProLysProLeuLeuCysSerGlnTy 2510
 Db 4899 CCTCGTGGCCCCCAACCAAGCTGGGAGCAGAGAGCCCAAGCCACTGCTGTGCGAGTA 4958
 QY 2510 rGluThrLeuSerAspSerGlu 2517
 Db 4959 CGAGACACTCTCCGACAGCGAG 4980
 RESULT 9
 S83390 2930 bp mRNA linear PRI 28-MAR-1997
 LOCUS T3 receptor-associated cofactor-1 [human, fetal liver, mRNA, 2930
 DEFINITION nt].
 ACCESSION S83390
 VERSION S83390.1 GI:1911769
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2930)
 AUTHORS Sande,S. and Privalsky,M.L.
 TITLE Identification of TRACs (T3 receptor-associated cofactors), a
 family of cofactors that associate with, and modulate the activity
 of, nuclear hormone receptors
 JOURNAL Mol. Endocrinol. 10 (7), 813-825 (1996)
 MEDLINE 96408715
 PUBMED 8813722
 REMARK Genbank staff at the National Library of Medicine created this

entry [NCBI gibseq 179975] from the original journal article.

FEATURES
source

Location/Qualifiers
1..2930
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

gene

1..2930
/gene="TJ receptor-associated cofactor-1, TRAC-1"
456..2765

CDS

/gene="TJ receptor-associated cofactor-1, TRAC-1"
/note="Authors report detecting several TRAC alternative transcripts; TRAC-1"
/codon_start=1
/product="TJ receptor-associated cofactor-1"
/protein_id="AA050847.1"
/db_xref="GI:1911770"
/translation="MAQRADMLRGLSRESLALNTVAAGRGITIDLSQVPHLPVLP
TPGTPATAMORLATLPAPQPFSSRHSSPLSPGPTHLTPPTTSSERDRDR
DRDRERKSLITSTTVVHAPIWRPTEQSGSGSGSGSGSSRASHSHOHSP
ISPTQALQORPSVLNHTGKGIITAVESTPVLSTSTSPVRPAATPPTHCP
LGGLDGVYPTLMEPVLLPKRAPVARPERPADTGAFLAKPPRAGSLPASPSSG
SEPRLYPVSGHATTARTPAKNTAHHASDPDPAAPASADPHRETKSPESDGL
ELSLGHHGSSSPSEGVSPVSPSSPDLTDKGLPKLEEDKSHLEGEIRPKQPGV
KLGGEAHLPHLRLPESQSSPDLQTAQVKGHQVVTLAQHI SEVITQDTRHP
QQLSAPLPAPLYSPFGASCPVLDLRPSDLYLPDPHGA PARSGPSEGKSPBN
KTSVLGGEGDIEVSPDEGTEPGRHSRAVYPLLYDGEOTBSRMSKSPGNTSOP
PAFSPKLTESNIAVSKQKINKLNTNRNPEVNI SODPTEIFMPPIITGGMT
YRSQAVQEHASTNGLEAITRKALMGCGAKKSGRSSKASPAAGLASGDRPSV
SSVHSEDDCNRRTPLTNRWEDRPSAGSTPFPYNPLMLQGVMASSPPPLPAGS
GPLAGAHAWDEERKPLCSQYETLSDSE"

ORIGIN

Alignment Scores:

Pred. No.: 2.05e-85 Length: 2930
Score: 4612.00 Matches: 873
Percent Similarity: 94.38% Conservative: 1
Best Local Similarity: 94.28% Mismatches: 6
Query Match: 34.90% Indels: 46
DB: Gaps: 1

US-09-522-753-5 (1-2517) x S83390 (1-2930)

QY 1592 ILeAlaLySeRProHISerThVaIProGluHISerProHISerProTyR 1611
Db 123 CTGGGCTCACCCCGCAGACCGTCCCGAGACACACCCCATCTGCCTAT 182
QY 1612 GLuHISleuLeuArgLyVaISerGlyVaIASerLeuTyraISerSerHISleProLeuAla 1631
Db 183 GAGCACCCTGCTCGGGCGGTGAGTGGAGCCTGTATGACGACCATCCCTGGCC 242
QY 1632 PhASpProHISerHISerProArgGlyLeProLeuSPrAlaAlaAlaTyTYrLeu 1651
Db 243 TTGACACCCACCTTCATACCCCGGCGATCCCTTGACGAGCGGTGCTTATACCTG 302
QY 1652 ProArgHISleuAlaProAsnProThTyRProHISleuTyRProTyTYrLeuAlaArg 1671
Db 303 CCGGACACCTGGGCCCCCAACCCACCTACCCGACCTGTATCCACCTCATATCCG 362
QY 1672 GLyTYrProAsnThraAlaAlaLeuGluAsnArgGluInThIleIleAsnAspTYrIleThr 1691
Db 363 GGGCTACCCCAACGCGCGGCTGGAGAACCGGACGACATCATCATCATCATCACCC 422
QY 1692 SerGluGluMetHISISerThraAlaThraAlaMetAlaGluAsnArgAlaAspMetLeuArg 1711
Db 423 TCGCAGACAGATGACCAACAGCGGCGACGACCTGACGAGCTGATATGCTGAGG 482
QY 1712 GLyLeuSerProArgGluSerSerLeuAlaLeuAsnTYrAlaAlaGlyProArgGlyIle 1731
Db 483 GGCCTGCGCGCGCGAGTCTCGCTGCACTCAACGTGCGGATCCCGGAGGATC 542
QY 1732 ILeAspLeuSerGluAlaProHISleuProValIleuValaProProThrProGlyThnPro 1751
Db 543 ATGACCTGTCACAAAGTGCACACTGCTGTGCTCGTGGCCCCGACACAGGACCCCA 602

QY 1752 ALaThraAlaMetAspArgLeuAlaTYrLeuProThraAlaProGluProPheSerSerArg 1771
Db 603 GCCACCCGCTGACCGGCTTGGCTTACTCCACCGCGGCCGCTTTCAGACGCCG 662
QY 1772 HIsSerSerSerProLeuSerProGlyGlyProThrHISleuThryIleProThrThn 1791
Db 663 CACAGCAGCTCCCACTCTCCGAGAGTGCACACACTTGAACAAACCAACCAACAG 722
QY 1792 SerSerSerGluArgGluArgAspArgGluArgAspArgAspArgAspArgGlu 1811
Db 723 TCTCTGTCGAGCGGAGCGAGACCGGATCGAGAGGAGGACCGGATCGGAGCGGAA 782
QY 1812 LySerHISleuThrSerThrThnThraValGluHISAlaProIleTPArgProGlyThr 1831
Db 783 AAGTCACTCTTCACTTCACTACGAGCGGTGAGACGACCATCTGAGACCTGTACA 842
QY 1832 GLuGluSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgPro 1851
Db 843 GAGCAGAGCAGCGGACAGCAGCGAGCGCGGGGTGGGGGACAGACAGCGCGCCC 902
QY 1852 AlASerHISerHISAlaHISGluHISerProIleSerProArgThrGluAspAlaLeu 1871
Db 903 GCCTCCCACTCCCATGCGCACAGCACTCGCCATCTCCCTCGGACCCAGATGCCCTC 962
QY 1872 GLuGluArgProSerValIleuHISerThryIleIleThraAlaValGlu 1891
Db 963 CAGCAGAGACCCAGTGTGCTTCAACACAGCATGAAAGGTATCATACCGCTGTGGAG 1022
QY 1892 ProSerLySerProThraValIleuArgSerThnSerSerProValaArgProAlaAla 1911
Db 1023 CCCAGACGCGCCAGCGTCTGAGTCACTCCACCTCCACCGGTGCGGACGCTGCC 1082
QY 1912 ThnProProAlaThnHISCyProLeuGlyGlyThryIleuAspGlyValTYrProThr 1931
Db 1083 ACATTCCTCCACTCCACCTCCACCTGCGGCGGACCTTCATGCGGCTTCACTACC 1142
QY 1932 LeuMetGluProValIleuProLyGluAlaProArgValaAlaArgProGluArgPro 1951
Db 1143 CTATGAGCCGCTTCTGCTGCTCCAGAGAGGCCCGCGGATCCCGGACAGCGGCC 1202
QY 1952 ArgAlaAspThryIleHISAlaPheLeuAlaLyProProAlaArgSerGlyLeuGluPro 1971
Db 1203 CAGCAGAGACCGGCGATGCTTCTCGGCAAGCCCGGCTCGGGGTGGAGCCC 1262
QY 1972 AlASerSerProSerLySerGlySerGluProArgProLeuValaProProValaISerGlyHIS 1991
Db 1263 GCCTCTCCCGCCAGCAAGGCGTGGAGCCCGGCTTGTGCTCTGCTGGCCAC 1322
QY 1992 ALaThryIleAlaArgThrProAlaLyAsnLeuAlaProHISAlaSerProAspPro 2011
Db 1323 GCCACATGCGCGGACCCCTCGGAGAACCTTGACCTTCACACGCGGACCGGAGCCG 1382
QY 2012 ProAlaProProAlaSerAlaSerAspProHISArgGlyLySerThryIleSerLyProPhe 2031
Db 1383 CCGGCGGACCTGCTCGGCTCGGACCGGCAACGGGAAAGACTCAAGTAACCTTT 1442
QY 2032 SerIleGluGluLeuGluLeuArgSerLeuGlyTYrHISLySerSerTYrSerProGlu 2051
Db 1443 TCCATCAGAGACTGGAATCCCGTCTCGGGTTACACGCGGACGACTACACCCCGGAA 1502
QY 2052 GLyValGluProAlaSerProValaSerSerProSerLeuThnHISAspLyGlyLeuPro 2071
Db 1503 GGGGTGAGCCGCTTCAAGCCCTGTGACTTACCAACAGTCTAACCACAGAAAGGGCTCC 1562
QY 2072 LyHISleuGluGluLeuAspLySerHISleuGluGlyGluLeuArgProLyGluPro 2091
Db 1563 AAGCAGCTGAGAGAGCTTCAGCAAGACCACTGGAAGGGGAGGTGCGGCCAAGAGCCA 1622
QY 2092 GLyProValaLyLeuGlyGlyGlyAlaAlaHISleuProHISleuArgProLeuProGlu 2111
Db 1623 GGGCGGTGAAAGCTTGGCGGAGGCGGCCCACTTCCACACTGCGGCGGCTGCTGAG 1682

QY 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValIleGlySerGlnArg 2131
 DB 1683 ACCGAGCCCTCGATCCAGCCGCTGCTCCAGACCGCCCGAGGGGTCAAAGTCAACGCGCG 1742
 QY 2132 ValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHis 2151
 DB 1743 GTGGTCACCTGGCCCGACATCAGTGAATCATCACAGACTACACCCGCGCACAC 1802
 QY 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCys 2171
 DB 1803 CCACGACGACTCAGCGCACCTCCGCGCCGCCCTCTACTCTCTCTGGGCGAGCTCC 1862
 QY 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGly 2191
 DB 1863 CCGGCTCTGACCTCCGCGCCGACCCAGTGAATCTTACTCCCGCGCCGAGCATGTGT 1922
 QY 2192 AlaProAlaArgGlySerProHisSerGluGlyValArgArgSerProGluProAsnLys 2211
 DB 1923 GCGCCGCGCCGCTGGCTCCCCCAGACGAGGGGCGMAAGGTCTCCAGAGCCAAACAG 1982
 QY 2212 ThrSerValLeuGlyValGlyValAspGlyIleGluProValSerProGluGlyMet 2231
 DB 1983 AGCTCGGCTTGGGTGTGTGTGAAGACGTTATTGAACCTGTGTCCCGACCGAGGCGATG 2042
 QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
 DB 2043 ACGGAGCCAGGCGACTCCCGAGTGTGTGTACCCGCTCTGTACCGGAGTGGGGAACAG 2102
 QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
 DB 2103 ACGGAGCCAGAGATGGGTCTCCAAAGTCTCCAGGACACACGACGCGCGAGCTTTC 2162
 QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsn 2291
 DB 2163 TTCAGCACTGATCCAGACGACCACTCCGCTCATGTCAAGTCCAGAGACCAAGATCAAC 2222
 QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThr 2311
 DB 2223 AAGAAGCTGAACCCACACCGGAATGAGCTGATCAATCAATTCAGCCAGCTGGGAGCG 2282
 QY 2312 GluIlePheAsnMetProAlaIleThrGlyTyrThrGlyLeuMetThrTyrArgSerGlnAla 2331
 DB 2283 GAGATCTTCAATATGCCGCCCATCACCGGAAACGCGGCTTATGACCTATGAAGCCGAGGG 2342
 QY 2332 ValGlnGluHisIleAspThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
 DB 2343 GTGCAGGAACATGCCAGACCAACATGGGGCTGAGGCCATTAATTGAAAGGACACTCATG 2402
 QY 2352 GlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
 DB 2402 ----- 2402
 QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
 DB 2402 ----- 2402
 QY 2392 HisThrLeuThrSerProGlyValGlyValLysAlaLysValSerGlyArgProSerSer 2411
 DB 2403 -----GTTGGCGGCGGAGGAGCCAAAGTCTCTGGCAGACCCAGCAGC 2444
 QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
 DB 2445 CGAAAGCCCAAGTCCCCGGGCGCGGCTGTGGCATCTGGGAGCCGCGCACCTGTCTCC 2504
 QY 2432 SerValHisSerGluGlyAspCysAsnAlaGlyArgThrProLeuThrAsnArgValTyrGlu 2451
 DB 2505 TCAGTGCACCTCGGAGGAGATGCAACCGCGGACCGCTCTCAACACCGCGTGTGGAG 2564
 QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLys 2471
 DB 2565 GACAGGCCCTCGTCCGACAGGTTCCAGCCATTCCTCTCAACCCCTGTGATCATGCGGCTG 2624
 QY 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491

DB 2625 CAGGCGGGGTGTCATGCTTCCCAACCCCGGCGCTCCCGCGGAGCGGCGCCCTC 2684
 QY 2492 AAGlyProHisIleAlaTyrAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
 DB 2685 GCTGGCGCCACACGACCTGGGAGCAGAGGCCCAAGCCACTGCTGTCTCGAGTACAG 2744
 QY 2512 ThrLeuSerAspSerGlu 2517
 DB 2745 AACTCTCCGACAGCGAG 2762
 RESULT 10
 AB028970 7949 bp mRNA linear PRI 02-APR-2003
 LOCUS Homo sapiens mRNA for KIAA1047 protein, partial cds.
 AB028970
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 Kikuno, R., Nagase, T., Ishikawa, K., Hirose, M., Miyajima, N.,
 Tanaka, A., Kotani, H., Nomura, N., and Ohara, O.
 Prediction of the coding sequences of unidentified human genes.
 XIV. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro
 DNA Res. 6 (3), 197-205 (1999)
 MEDLINE
 PUBMED
 99397452
 10470851
 2 (bases 1 to 7949)
 REFERENCE
 AUTHORS
 Ohara, O., Nagase, T., and Kikuno, R.
 Direct Submission
 Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913,
 Fax: +81-438-52-3914)
 On Mar 31, 2003 this sequence version replaced gi:5689430.
 COMMENT
 FEATURES
 source
 1..7949
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /clone="h01221s1"
 /tissue_type="brain"
 /note="vector: pBC SK plus"
 1..7949
 /gene="KIAA1047"
 <12..7058
 /gene="KIAA1047"
 /note="Start codon is not identified.
 This sequence was obtained by subcloning of the DNA
 fragments derived from two cDNA clones (1 - 1691 was
 derived from h01221 and 1692 - 7949 was derived from
 h02672)."

/product="KIAA1047 protein"
 /protein_id="BA02999.2"
 /db_xref="GI:29421188"
 /translation="SHSLVRTSLIMSSGYPPNGAFSTEQSRYPHISVOYTPNTR
 HQEFAVADYSSHLEVOASQLQOQOQOQRRPSLSSEHPGSDRPOERTSYEP
 FHPGSPVHDLSLEKRPRLCOVSDSHQVSAVLPVHPLEGLASADAKQDAF
 GSKHAPSPSISGPGCDONASPSKLSKEELISGMDVDRILAVEQOILKTKKQO
 QLEEAARPREPRKVPSPPYQKRSIVQIYIDNRKKABEAKHIFSGIGPKVPLP
 INQPSDVTYHENIKTNQWRKGLIFIKRNNHAKQBPQKICQYTDLMEMWEKVD
 RIENPRRKAESKTRREYKQPEIRKQREQERFQVGGAGLSATIASSEHS
 EIIIDLSQENNEKMRQLSVLPMPFPAEGRVFEINNGIMEDPMKRYKORPMNV
 WTDHEKEIFQKFIQHPKNGLIASYLRKSKVDCVLYYILTKKENYALVRRYK
 RRGNOQIARPSQEKVEKEDEKAKETKEKEKDEKDEKDEKDEKDEKDEKID
 GTAETREQATTPRGRTANSQGRKRKIRTSRWNTEAASAAASAAATEEPPLPP
 PPEPISTEPVETSRWTEEMEVAKGLVHEGRNMAIALIKMTKSEACCKNFYNYKR

[illegible]

Db	2547	-----GCTGACTCTGTGAGACCTTGGAAGTAGCGGTGCAGAAAACCAATGCATCTAAAGTT	2600
Qy	871	GlutGluGlyProAlaAlaValGlyValAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla	890
Db	2601	GAAGGTGATATATACCAAGAAAGAGAGCTTGATGAGCC-----	2639
Qy	891	LeuValAlaGluValValGlyGlySerGlyValThrThrAlaValSerSerGly	910
Db	2640	-----AGTGAAGAGGTGGAACTTAGAGATGAGAGATTGTTGGTGAAGCTCAGCAATTAAT	2693
Qy	911	Ala-----ProGln-----AspSerAspSerSerAlaThrCysSerAlaAspGlu	925
Db	2694	GCCCAAGAGCCCGAGCCCGCCAGCTCAGACAAATATTCCAGTGCACGTGCAGCGGTGATGAG	2753
Qy	926	ValAspGluAlaGluGlyGlyAspValAspAsnArgLeuLeuSerPro-----ArgPro	942
Db	2754	-----GATGGATGGAGGAGCCAGAGGACAGAGAGAGATGTTTCCATAGAGACTCAAAAGCCT	2807
Qy	943	SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----Pro	960
Db	2808	TCACCTGTTAAACCCACCTGATCTTAACTGCTC--TCATCTCCGTTAAACCAATCCA	2864
Qy	961	LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaAlaIleProPheIle-----	976
Db	2865	CTGATCTGCCACAGAGCTTCACAGATCAGAGCTGCTTTATCCACCAATGATCTCTGCACC	2924
Qy	977	-----GlnVal	978
Db	2925	CCATGTAACATACCAATTGGAACCCAGTCAGCGCTATGCTCTTACAGCGACACATT	2984
Qy	979	ThrIysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAla	998
Db	2985	AAAGCAATGATGAGTGCAGACTCTCTGGAG-----	3014
Qy	999	ProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer	1018
Db	3015	-----GAGCAGCGCGCAGACAGCAAGAACAGATATATTGGAATGATGAAGTTCACAGT	3068
Qy	1019	ProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGlu	1038
Db	3069	CCATGTGGCACATCCAGAGTCCA-----AACAGAGAG-----	3101
Qy	1039	AlaGlnLysLeuProGlyAspProProGlySerPheThrSerGlyLeuProPheProAlaPro	1058
Db	3102	-----TGG-----	3104
Qy	1059	ProArgGluValIleIleValSerProHisAlaProAspProSerAlaPheSerTyrAla	1078
Db	3105	-----GAGTCTCTTCAAGCTCTCCACAT-----CAAGATATACTTAATCTC	3146
Qy	1079	ProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg	1098
Db	3147	CCTGAAGCGTTCGGGCTTCG-----ACAACCTGACCAACAGCGCCACCG	3191
Qy	1099	ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeu	1118
Db	3192	CCCCCTCTCATCCCGCTCATCCAAAACCAAGTGGCTTCAGAAAAA---CCATCTTTTATA	3248
Qy	1119	GluArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeu	1133
Db	3249	-----ATGGAGAGGCTCATCTTCAAGGAAACACCAGGACCTATTGACCTTCTCATAT	3302
Qy	1134	HisValProTyrSerGluHisAlaLysAlaPro---ValGlyProValThrMetGlyLeu	1152
Db	3303	CAGGTTCTCTACCTCAAGAAACACCCAGCGTCAGTGGGATCTATCTCTTGGAGAT	3362
Qy	1153	ProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnGlnLeu	1172
Db	3363	CCACGGCAACAGGATGCGCAAAATGACCTACTTGGCTTCATCAACAGAGAAAGAAATT	3422
Qy	1173	SerProAlaArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAla	1192
Db	3423	TCTCCCGAGAACCAAACTCAAACTGATGGGCTCTGTTGGTC---AGGGCCCAACATGAA	3479

Oy	1193	SerValLeuAlaGlyThrAlaLeuGlySerValProGlyGlySerTLeThLyGlyTLe	1212
Db	3460	GGTATGTCAGAGGATACCGCA---GGACCATCAAGAAAGATATATCTCGGGAACT	3536
Oy	1213	ProSerThrArgValProSerAspSerAlaLeThTrYrArgGlySerTLeThHieGly	1232
Db	3537	CCACACGAAATTTCTAGTGGAGAGCATTCATCTCAAGGGCTCTATCATCAGGGC	3596
Oy	1233	ThrProAla-----AspValLeuTrYrLyGlyTLeTh	1244
Db	3597	ACCCGGCTCGCCCCAGACTGGCATACCAACAGAGGCTTTGGTGAAGGGGTCATTTCG	3656
Oy	1245	ArgTLeileGlyGluAspSerProSerArgLeuAspArgGlyValGluAspSerLeuPro	1264
Db	3657	AGAAATGCCCATTTGAAGACAGCATGCTCT-----GAGAAAGCCAGAGAGAAAGCTGCATCC	3710
Oy	1265	LyseGlyHsValLileTrGlyGlyLysValysGlyHsValLeuSerTrYrGluGlyGlyMet	1284
Db	3711	AAAGGCCATGTTATTTATGAAGGCAAAAGTGACATATCTGTCTATATGAATATTTAAG	3770
Oy	1285	SerValThrGlyCysSerLyseGlyAspGlyArgSerSerSerGlyProProHieGluThr	1304
Db	3771	AATGCC-----CGAAGACGACTAGACTCCAGAAACGCTCATGAATATC	3815
Oy	1305	AlaAlaProLybArgThTrYrAspMetMetGlyArgValGlyArgAlaIleSer--	1323
Db	3816	AGTTTA---AAGAGAACCTATGAATCATGTCAGAAAGGAATATTAAGCAAGGATGTCATG	3872
Oy	1324	-----SerAlaSerTLeGlyGlyLeuMetGlyArgAlaIleProProGlu	1338
Db	3873	AGGAGTCTCTGATTCAGACACCGTTAAGGGGCTGATATGCGCGAGCTTACCC-----	3926
Oy	1339	ArgHieSerProHieS---HieLeuLyseGlyGluHieHieSileArgGlySerTLeThGln	1357
Db	3927	AGGGGAGTCTCTCATTCGTACACTCAAAAGAAAGACTGATTTGTCTGCTCATATAGAG	3986
Oy	1358	GlyTLeProArgSerTrValGluAlaGlnGluAspTrYrLeuArgArgGluAlaLyLeu	1377
Db	3987	GGGACACCAAGCAACCACTGAAGAGCTTTAAGATGAGCTTT---AAATATCCAAACAA	4043
Oy	1378	LeuLybArgGlyGlyThrProProProProProSerArgAspLeuThGluAlaTrx	1397
Db	4044	ATTAAAGGAGAAAGT-----CTCCCATACAGACATTTGAAGGTGCCATT	4088
Oy	1398	LyThrGlnAlaLeuGlyProLeuLyLeuLybProAlaHieGlyGlyLeuValAlaThr	1417
Db	4089	-----ACCAAGAAACCA---TATGATGCG---ATCACACCC	4121
Oy	1418	ValLyseGlyAlaGlyArgSerTLeHieGlyLleProArgGlyGluLeu-----	1433
Db	4122	ATCAAAAGAAATAGGGCGCTTTCATTCATGATGATTCACAGCAAGATATTTTAACTCAGAA	4181
Oy	1434	---ArgHieThrProGluLeuProLeuAlaProArgProLeuLybGlyGlySerTLeThr	1452
Db	4182	AGTCGGAATCTCCAGAGGTGCTCAGAGCACAGCGCGGATATTAAGGGTTCCATTTC	4241
Oy	1453	GlnGlyThrProLeuLyTrYrAspThrGlyAlaSerThrThrGlySerTrYrLybHieAsp	1472
Db	4242	CAGGCAACACAAATTAAGTTTGACAACAAC---TCAGTCAATCTGCCATCAACACAAAT	4298
Oy	1473	ValArgSerLeuLileGlySerProGlyArgThrPheProProValHieProLeuAspVal	1492
Db	4299	GTCAAAATCTTAAATCACGGGGCTTACCAAACTATCCCGTGAATAGCTCCGCTGAAATTT	4358
Oy	1493	MetAlaAsp---AlaArgAlaLeuGlybArgLalaCyTrYrGlu-----Glu	1506
Db	4359	GTCCCAAGAAACATTAAGGTGTACACCGGGAAATTTTGAGATGTGAACAGCGGAG	4418
Oy	1507	SerLeuLybSerArgProGlyThrAlaSerSerSerGlyGlySerTLeAlaArgGlyAla	1526
Db	4419	ACCGGACGTTCCCGGACACGTCAGTGTAAAGTCTGGCCCTCGCTTAAAGTTCACA	4478

QY	1527	ProValIleValProGluLeuGlyIleuProArgIleuSerProLeuThrTrpGluAspHis	1546
Db	4479	-----CGCATGAAGCTCCAAAGCAACAGCTGAGCTGGATTTAATGAAGACCC	4529
QY	1547	GIyAla-----ProPheIaGlyIleuProArgGlySerProValThr	1561
Db	4530	AGTGCACGAGGAGCCCTGTGAGTTATCAAAACACATGTCACAGGTCACCATG---	4586
QY	1562	MetArgIuProThrProArgLeuGluGlySerLeuSerSerIleValA---Ser	1580
Db	4587	-----ATGAACAGAACTTCGATGTTCAAAATTCCTTTACACAGCTCAACAT	4634
QY	1581	GIAspArgIleLeuThrSerThrThrProArgGlu-----IleAlaIleSerPro	1596
Db	4635	CATGAAGGAATATGACACATGACCCCTTACCCAGAGGAAAGTATCCAGCAAGTCTCA	4694
QY	1597	HisSerThrValProGluIleAspProIleSerProTyrGluIleuLeuArg	1616
Db	4695	GTGCTGGGGGTGACCTGTGTCGAGCAC-----AGTCGTTGATCCCATACACGA	4748
QY	1617	GIyAlaSerGIyAlaAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer	1636
Db	4749	GAGCAGCACTGCAGGCGAGGTTTATCGAGCCACCTGCCACGACTTGAAATCA---GCC	4805
QY	1637	IleProArgGlyIleProLeuAsp---AlaAlaIleAlaTyrTyrLeuProArgHisIleu	1655
Db	4806	ATGCTCTTTCACAGGCTTTGATCCCTGACGCGCTCTTACCTGTTCAAGACAGCTT	4865
QY	1656	AlaProAspProThrTyrProHisIleuTyrProProTyrIleuIleArgGlyTyrProAsp	1675
Db	4866	TCACCAACTCCAGGTTACCCAAAGTCAGATCAGCTTAC-----	4904
QY	1676	ThrIleAlaIleGluAsp---ArgGluThrIleIleAspArgTyrIleThrSerGluIle	1694
Db	4905	-----GCAATGGAGAAACACAGACAGACATCTTAATGATTTACATTACTCCACAAAG	4958
QY	1695	MetHisIleAspThrAlaThrAlaMetAlaGluIleAspMetLeuArgGlyLeuSer	1714
Db	4959	ATGCAGATGAACTTG-----CGTCAAGATGCGCAGACAGACTCTCC	5000
QY	1715	ProArgGlySerSerLeuAlaIleAspTyrAlaIleArgIleProArgGlyIleIleAspLeu	1734
Db	5001	CCAAAGAGCAGCCACCTGGCTGCTCCATACCACCA---ACGAGAGAAATATATGACCTG	5057
QY	1735	SerGluValProHisIleuProValIleuValProProThrProGlyThrProAlaThrAla	1754
Db	5058	ACCAATATGCTTCA---ACAATTTATGCTCATATCCAGGGGGAACAAGACTCTCC	5114
QY	1755	MetAspArgLeuAlaTyrLeuProThrAlaProGluProPheSerSerArg---HisSer	1773
Db	5115	ATGACACAGAACTACCTTATATCTCGTAGACACAGATTACTTCCCTCCAGGCGCTCAAC	5174
QY	1774	SerSerProIleuSerProGlyGlyIleProThrHisIleuThrIleuSerProThrThrSerSer	1793
Db	5175	TCTCTTCATAGTCTCCAGACACCAACACACTT-----GCACTGCTGCAGAT	5225
QY	1794	SerGIuArgIleuArgAspArgAspArgGlyIleuAspArgAspArgGlyIleuGlySer	1813
Db	5226	GCTGAGAGGGAGACGGGAACGGGAGACGGGAGAGAGCGGAGCGGAACGG-----	5276
QY	1814	IleLeuThrSerThrThrThrValGluHisAlaProIleThrArgProGlyThrGluGluIle	1833
Db	5277	ATTGCTCAGAGTTCCTCCGACCTC-----TACCTGCGGCGAGGCTCGAACAAG	5324
QY	1834	SerSerGlySerSerGlySerSerGlyGlyIleuGlyIleuSerSerSerArgProAlaSer	1853
Db	5325	-----CCTGGCCGACCTGGCAGT	5342
QY	1854	HisSerHisAlaHisGluHisSerProIleSerProArgThrGluAspAla---LeuGluIle	1872
Db	5343	CATGAGATATGTCCTCCCTTCCCTC---TCAATGAACATCAGGAGACCATGTTGCCAA	5399
QY	1873	GluArgProSerValIleuHisAspThrGluMetLeuArgGlyIleIleThrAlaValGluPro	1892

[illegible]

Db	6138	AAATCCCAAGAGAGAGTCAAGTC-----TCTTCGAGGCGCTTACAGACCATCTCC	6186
QY	2227	ProProgluGlYmeTThrGluProGluYHIsSerArgSerAlaValTyrProLeuLeuTyr	2246
Db	6189	CCACCCCCAG-----GTTCGCGTTGTGCATGAAACAGACACAGCTGGCTGCTTGTCT	6242
QY	2247	ArgAspGluYgluGlnThnGluProSerArgMetGlySerIySerProGluYAsnThrSer	2266
Db	6243	CAGAGGGGCGCAGAGCTCGCAGACAGAGAAATATATGCCCTCACCGAGGAGTAAAGC	6302
QY	2267	GlnProProAlaPhePheSerIyLeuThrGluSerAnsSerAlaMetValIySerIyS	2286
Db	6303	TACTTGGCTTCATCTTCCACCAAGCTT---GAAATATCATCACCACATGGTAAATCAAA	6359
QY	2287	LysGlnGluIuIeAsnIyIyIyLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle	2306
Db	6360	AAGCAGGAGATTTTTTCGTAAGTTAACTCTCTGTGAGAGGTGACCTCGAATGGACGT	6419
QY	2307	SerGlnProGluYHnGluIlePheAsnMetProAlaIleThrGluYHnGluLeuMetThr	2326
Db	6420	GCTACGACGAGAACTCGAGATCTTAACTGTGCACAGTACTACGTACGAGCTTCAGTTAG	6479
QY	2327	TyrArgSerGlnAlaValGlnGluIuIeAlaSerThrAsnMetGlyLeuGluAlaIleIle	2346
Db	6480	TCTAGAGGCCATCTCTTTGTCGATCCGACGCT---AATCTGGCGTGGAAAGACATTATC	6536
QY	2347	ArgIySalAlaLeuMetGlyIyIyTyrAspGlnTyrGluGlu-----Ser	2360
Db	6537	AGGAAGGCTCTCAATGGAGAGCTTTGATGACAAAGTTAGATCATGAGATTGATGATGCC	6596
QY	2361	ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaIa	2380
Db	6597	CAGCCTAATGGAGTA-----GTGCTTGATCTGCTGCACACCTCATGTT-----	6638
QY	2381	MetProIleThrAlaAlaAspGlyIyArgSerAspHisThrLeuThrSerProGlyIyGly	2400
Db	6639	-----GTGACCAATGGTGAGACACAGAAAGAGAGAGAGAGGCCCATCATCTCATTCAGA	6692
QY	2401	GlyIy-----IySalAlyIyValSerGlyIyArgProSerSerArgIySalAlyIySerPro	2418
Db	6693	GGAGTTTGCAAAACCAAGCTGATCAGCAAGTCAACCAAGCAGGAATCATTAATCTCTCAT	6752
QY	2419	Pro-----GlyLeuAlaSerGlyIyAspArgProSerIyAlaSerIyValHisSerGlu	2436
Db	6753	CCTGGGCAAGGCTACTTAGAAGCGAAGCGCCTCTTCAGTCTCTCTGTAATTCAGAA	6812
QY	2437	GlyIyAspCyAsnArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSer	2456
Db	6813	GGGATTAACCTAAGGCACAGCCCA-----GGTGGGCGCTGGGAAGACAGGCCCTCTTCA	6866
QY	2457	AlaGlySerThrProPheProIyTyrAsnProLeuIleMetArgLeuGlnAlaGlyIyMet	2476
Db	6867	ACAGGCTCAACTCAGTTTCTTATAACCTGTACTGATGCGGAGT-----CTCAGC	6917
QY	2477	AlaSerProProProProGluYLeuProAlaGlySerGly---ProLeuAlaGlyProHis	2495
Db	6918	AGTACTCCACCAACACCGATGTGATGTGCTCCCTCCGCGTAAACAAGCAGCTCTCAAC	6977
QY	2496	-----HisAlaTyrAspArgGluGluProIyAspProLeuIyCySerGlnTyrGluThr	2512
Db	6978	CAACAGAACAGAGATCTGGAGCGAGAGCTCCCCCATCTGCTCAGCAGATACAGACAC	7037
QY	2513	LeuSerAspSerGlu 2517	
Db	7038	CTGTCCGATTAAGT 7052	
RESULT 11			
MMU35312			
LOCUS			
DEFINITION Mus musculus nuclear receptor co-repressor mRNA, complete cds.			
ACCESSION U35312			
VERSION U35312.1 GI:1022717			
KEYWORDS			

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 7780)
AUTHORS Horlein, A.J., Naeir, A.M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Kama, Y., Ryan, A., Glass, C.K. and Rosenfeld, M.G.
TITLE Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor
JOURNAL Nature 377 (6548), 397-404 (1995)
MEDLINE 96008639
PUBMED 7566114
REFERENCE 2 (bases 1 to 7780)
AUTHORS Horlein, A.J., Naeir, A.M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Kama, Y., Ryan, A., Glass, C.K. and Rosenfeld, M.G.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1995) Andreas J. Horlein, School of Medicine, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0648, USA
FEATURES
source
1..7780
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/issue_type="pituitary"
117..7478
/note="N-COR"
/product="nuclear receptor co-repressor"
/protein_id="AA017125.1"
/translation="MSSSGYPNQAGFSTQSRSHSVQYTPFAPAHQOEPAVDK
SLSKPRLQVDSHQRISAVALPHLPGELSGSRDQRSGADPGVGHGPA
SGPCDDQNAHSPKSKSEKELISMDRYDEIAKVEQILKAKKQOELEAKPPE
PERVSPPEVQKHSTVQILYDENRKAEBAKHIEGLDLPKELPLVNPSTKYH
ENIKTNQWAKKILFPRKRNHAKOEQICQYDQLMKEXKVDRIINRRKX
ESKREYELQPEIRKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOE
NEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOEKQOE
KFIQHPKNEGLIASYLERKSPVCLVYVITKKNENKALVRENYGKRRNOIAP
SQEKYBEKEDAKETKEBEKEDKEDKEDKEDKEDKEDKEDKEDKEDKEDKED
VTGRKTAASQGRGKRVTRSTSEAAANAAATBEPPPLPEPPELSTEBET
SWTEEMEVAKKGLVEHGNMAIAVQTKBAQCNTKYFKYKRNHNDLQOHK
OKASRRPREEDVQCEVASVTAQDEDEIASENEENDEDEGANSDDTSASP
SPYBAKSSSESDENASRGTEPVAELAKTTPAPCASPSVPTKPEREASR
VDSASAEKTAEMDNDHEGCGSSILDPAPTKADSDPNQVVENTASKEGAK
ERLESTSEKTEKDEVDVAQIERPEQSDSDSATGADGVDEPERQVFPMD
AKSLTLPPGSLISSPIKPNLDLPOLQRAAVIPWVSCENPIPIGPVSGVALY
ORHIAKHESALIEBOROROEVDLECRSSTSPSTSKSPREMEVLQAPHOVITL
PEGVRLPTTPRPPLPSSKTVAEKSPTMGSGISQGTGYLSHNOVAPOE
APGVSGTSLGPROESTKAPLTYIKQEEFSPSONSGPGLVLRAGHEVAGT
AGAVGSGTGTCTPASTISVETISSLRGSGTCTPAPLPOGITEALYKGPVEMPIE
ESSEKVEKBEAKGHTVIGKSGHLSYDINIKNABGTPPTAHMSLKRSEAVE
GSTKQMSKESVSAVLEGLICRALPRGPHDLKERTVLSIGQGTGYLSHNOVAPOE
DGAKYKQIKRESPIRAGATGKGPYGTITIKEMGSHIEIPRODILTESEKRT
PEVQGTIRLIESISQGTPIKPDNNGQALKNVSLTGKSLPRGLMEIVPNT
KYVREGVEDVGAEPYRAHRTSVSSGSHVSLTLHAKALSGLVNDSABRP
VYQNTISGSGPMNRTSDVSSSKASHKSTLTPQKRSITAKSPREVDVDSHS
PPPHRRSSAGGVTRSHLTPHDPAPFRALDPAAVYLQQLSTPEYPSQYQL
AMNTQTLINDYTSQOMQVNLPRVTRLSPREQLPGPYATGAILDLTMRPPI
LVHAGGTTPMPDIRITVPGQVTPPPRYNAASLSPGHTALAAASGHEKREER
EKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKRE
QITTPDPAHQVSOASTSTFTQSPALSTPVRTKTSRSGVSPSOQTYLHRRP
VSEBNTLVKSGRSGPKSPERSHISEPYVISPPOPAVHEKQDMLLSQGVDA
BOBDSRSGSTSYLPSFTKLESTPMVSKQEIPIKLSNGGSDMAAOPCTE
IRULPAVTTSAGVSSSHSPADPASNGLDITRKALMSGFDKVDKGVMSHPGI
MPSASTSYVTSKARDGSGSPHAGVCGPKLINSKSGSPIPGQSYLSTERS
SVSSVHSEGDYHRTQPGMAWEDRSSSTGSGTPYNPILTRMLSSSTPTQIACPSAIT

ORIGIN QAAPHOQRIMEREPAPLISAOYETLSDSD"
Alignment Scores: 1,68e-76 Length: 7780
Pred. No.: 4187.00 Matches: 1104
Score: 53.94% Conservative: 347
Percent Similarity: 41.04% Mismatches: 798
Best Local Similarity: 31.68% Indels: 442
Query Match: 10 Gaps: 104
US-09-522-753-5 (1-2517) x MM035312 (1-7780)
Qy 16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
Db 165 CAGAGTCGATCCTTCCGACCTGCTTCCAGTACACCTTCCAGGCGCTCCAGACAG 224
Qy 36 AspValGlyLeuLeuGlnTyrGln-----HisHisSerArgTyrAlaSerHisLeu 53
Db 225 GAATTTCAGTCCCTGACCTACCGTTCTTCATCTTGAAGTCAGTCAGGCGTCCAGCTC 284
Qy 54 SerProGlySerIleIleGlnProGlnArgArgProSerIleLeuSerGluPheGln 73
Db 285 TTCAGACGACGACGACGACGACGCTTCCGACAGCGCTTCCGATTTCCGAGTTTCAC 344
Qy 74 ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
Db 345 CCGGCTTCCGACGACGCGCCGACGAA-----AGAGAGAGTGAATACAGACAGCTCCACCG 398
Qy 94 GluLeuGlyLeuSerGluMetGluPheIleGluSerIleArgProArgLeuGluLeu 113
Db 399 GGCCTTCCCGGCTGACCATGCTGCTGAGATCCAAAGCGGCTGCTGAGACAGCTT 458
Qy 114 -ProArgProLeu-----LeuArgProSerProLeuLeuAlaThrGlyGlnPro 129
Db 459 TCCGACTCCCATCTCCAGCGCATCAGTCCGCTGCTCTTGTGTGACACG-CTGCC 517
Qy 129 AlaGly-----SerGluAspLeuThrIleAspArgSerLeuThrGlyLeuGln 146
Db 518 AGAAGACGACGATCTTCTCCCAATGTAAGAAGATCCGACATTTGGATCAACATGA 577
Qy 146 u--ProValSerProProSerProProHisThrAspProGluLeuValProPr 165
Db 578 AGCTCTTCTCTCTCTCTCTCTGAGGACGATCGAGATGATCAGAACTCTCACCTTC 637
Qy 165 AlaGluSerLeuGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMe 185
Db 638 AAAACGTCAAGAGAGAGCTGATACAGACATGATCGTGATCGAGAAATTCGAA 697
Qy 185 TValGluGlnGlnIleSerIleAspLeuIleValGlnGlnGlnGlnGlnGlnGln 205
Db 698 AGTGAACAGCAGATCTTAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
Qy 205 AlaProProGluProGluLeuLeuSerProProProIleGluSerHisIleArgSe 225
Db 758 TAAACCCGACGAGCTGAGAGAGCTGATGCTCTCTCCGCTGAGAGAGAGAGAGAG 817
Qy 225 rLeuValGlnIleIleTyrAspGluAsnArgGlyValGlnGlnGlnGlnGlnGln 245
Db 818 TATTGTCCAAAATCAATTAAGACAGAAATGGAAGAGAGAGAGAGAGAGAGAGAGAT 877
Qy 245 uGluGlyLeuGluProGlnValGlnLeuLeuProLeuTyrAsnGlnProSerAspThrArg 265
Db 878 TGAAGCTCTTGGCCCAAAAGTTGAATCCGCTCTCAACACGAGCTCAGATCCAAAGT 937
Qy 265 nTyrHisGluAsnIleLeuValLeuGlnAlaMetArgValLeuLeuIleLeuTyrPhe 285
Db 938 GTACCAAGAGAGATCAAG 997
Qy 285 sAlaGluAsnHisLeuAlaArgGlyGlnTrpGlyGlnIleSerGlyGlnArgTyrAspGln 305
Db 998 AAGAAGAAATCAATCAAG 1057

D	2999	AATTGGAAGCCCGTAAGTGGTAACTGCTTTACCAAGGCAATTAAGGCCATGATGA	3058
Q	983	uProBoArgIuaBpAlaAlaProThLyBProAlaBProProAlaProProProGI	10030
D	3059	GTCAAGCACTCTGGAG-----GAGCAGCG	3082
Q	1003	naBnLeuGIProGIuSerAaBpAlaProGIInPrProGIySerProAArgGlyLe	10233
D	3083	GCAGAGGCAGAACGCTGAGACTTGGAAATCGAAGACTCTCAAGCCCATGACACTTC	3142
Q	1023	rArgSerProAlaProProAlaBpLyGlyuAlaBpAlaAlaGIuAlaGIuLyLeuPr	10434
D	3143	TAAGAGTCA-----AACAGGAG-----	3161
Q	1043	oGlyAspProBoCySTrPThrSerGlyLeuProBheProValProProAArgIuAl	10633
D	3162	-----TGG-----GAAAGTCT	3172
Q	1063	eLyAlaSerProHlaAlaProAaBpSerSerAlaBheSerYrAlaProProGIyHlaPr	10833
D	3173	CCAACTGCTCGCAT-----CAAGTAAATCACTTCTGAAAGGGGTTCG	3220
Q	1083	oLeuProLeuGIyLeuHlaBpThrAlaArgProValLeuProAArgProThrIleSe	11030
D	3221	GCTTCA-----ACACACGACCAACACGCGCACCACTCTCCATCTCC	3265
Q	1103	rAsnProBoProBoLeuIleSerSerAlaLyHlaBpSerValLeuGIuAArgInIleGI	11233
D	3266	ATCATCTAAACCAACAGTGCATCGAATA-----CATCTTTATA-----ATGGAGG	3316
Q	1123	yAlaIleSerGIuGIy-----MetSerValGIuLeuHlaBpProTySe	1138
D	3317	GTCATCTCAACAGGAACTCTCGCACTTACTGTCTTCTATATACAGCTTATCCACA	3376
Q	1138	rGIuHlaLyAlaBpAlaProValGIyProValThMetGIyLeuProLeuProMetAaBpR	1158
D	3377	AGAAAGCCCTTAAGCCCTCGTGGGGTATATCTCTGGAAATGCCCGGACAGAGAGTC	3436
Q	1158	oLyAlaLyLeuAlaProBheSerGIyValLyGIuGIuGIuLeuSerProAArgGIuAl	1178
D	3437	TACCAAAAGCACTCTTTGACCTACACACAGGAAGAAATTTCTCCGAAACCAAA	3496
Q	1178	aGlyProBoGIuSerLeuGIyAlProThAlaGIuAlaSerValLeuAArgIyTh	1198
D	3497	CTCAAACTGAGGGTTATTATGGTC-----AGAGCCAGCATGAAAGTGTGCAAGGAC	3553
Q	1198	rAlaLeuGIySerValProGIyGIySerIleThLyGIyIleProSerThAArgValPr	1218
D	3554	TGCA---GGGGCTGTCAGAAAGAAAGTAACTCGGGGAATCTCAGCAACAAATCTC	3610
Q	1218	oSerAspSerAlaIleThrTyrArgGIySerIleThHlaGIyThProAla-----	1235
D	3611	AGTGGAGACATTTCATCTGTCGGGGGTCTATTATCCAGGGGACCCAGCTGCGCCA	3670
Q	1236	-----AspValLeuTyrLyGIyThrIleThrArgIle---IleGIyGI	1249
D	3671	GGCTGGAATACCAACAGAGCTTTGGTGAAGGACCTGTCTCCAGATGCTATTGAGA	3720
Q	1249	uAspSerProSerArgLeuBpAlaArgGIyArgGIuBpSerLeuProLyGIyHlaIle	1269
D	3731	AAGCAAGTCTTGAGAGGTC-----AGAGAGAAAGCTGCATCCAAAGGCCATGTAT	3781
Q	1269	eTyGIuGIyLyAlaLyGIyHlaIleSerTyGIuGIyIleMetSerValThrIncy	1289
D	3782	CTATGAAGCAAAAGTGCATATCTTATCTATGATTAATTTAAGATGCC-----	3833
Q	1289	BSerLyGIuBpGIyArgSerSerGIyProProHlaGIuThrAlaAlaProLySar	1309
D	3834	-----CGAAGAAAGGACTCGAGACTCCAAAGACAGCTCATGAATATGATTAA-----AAAAG	3883
Q	1309	gThrTyrAspMetMetGIuGIyArgValGIyArgAlaIleSer-----	1323
D	3884	AAGCTATGAGCAATGAAAGAAAGTATTAACAAAGCAAGCTGTGATGAGAGACTCTCTGT	3943

[illegible]

Oy	1020	ArgGlyValSerArgSerProAlaProProAlaSerGlyValAlaPheAlaAlaGluAla	1033
Db	3220	TGTGGCAATCCAAAGATGCCA-----AACAGAGAG-----	3249
Oy	1040	GlnGlyLeuProGlyAAspProProCysTrpThrSerGlyLeuProPheProValProPro	1055
Db	3250	-----TGG-----	3257
Oy	1060	ArgGluValIleGlyValAserProHlaAlaProAspProSerAlaPheSerTyrAlaPro	1079
Db	3253	---GAAATCCTTCAAGCTGTCACAT-----CAATTGATTAATTAATCTCCT	3297
Oy	1080	ProGlyHlaProLeuProLeuGlyLeuHlaAspTrpAlaArgProValLeuProArgPro	1099
Db	3298	GAAAGCGCTTCGGCTTCGG-----ACAACTCCGACCAACGGCCACCGGCC	3342
Oy	1100	ProThrIleSerAsnProProProLeuLeuIleSerAlaGlyHlaPheSerValLeuGlu	1119
Db	3343	CCTCTCAATCCGCTCATCCAAACCAACAGTGGCTTCGAAAAA--CAATCTTTTATA--	3366
Oy	1120	ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHla	1134
Db	3397	---ATGGAGAGGCTCATCTATCACAGGGAACACAGGCACTATTTCACCTTCATTAATAG	3453
Oy	1135	ValProTyrSerGlnHlaIleAlaLeuPro--ValGlyProValThrMetGlyLeuPro	1153
Db	3454	GCTTCCTCACTCAAGAAACACCCAAAGCGTCAGTAGATCTATCTTGAGACTGCCA	3513
Oy	1154	LeuProMetAspProTyrLeuValLeuAlaProPheSerGlyValIleGlnGlnLeuSer	1173
Db	3514	CGGCACAGGAATCTGCCAAATTCAGACTTTGCCCTACATCACAGAGAAATTTTCT	3573
Oy	1174	ProArgGlyGlnAlaGlyProProGlySerLeuGlyValProThrAlaGlnGluAlaSer	1193
Db	3574	CCCCGAAACCAAACTCAACACTGAGGCTGTGTTGGTC--AGGCCCCAACATGAAGT	3630
Oy	1194	ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrGlyGlyLeuPro	1213
Db	3631	GTAGTCAGAGGTACCGCA--GAGACCATCAAGAAAGAGATATACTCGGGGAATCCA	3687
Oy	1214	SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHlaGlyThr	1233
Db	3688	ACCAACCAAAATTCAGTGGAGAGCATTCATCCCTACCGGGCTCATCTACATCAAGGCC	3747
Oy	1234	ProAla-----AspValLeuTyrLeuGlyThrIleThrArg	1245
Db	3748	CCGGCTCTGCCCCAGACTGCGCATACCAAGAGCGTTTGGTAAGGGGTCCATTTCCGA	3807
Oy	1246	IleIleGlyGluAspSerProSerArgLeuAlaArgGlyArgGluAspSerLeuProlys	1265
Db	3808	ATGCCCATGTAAAGACAGATCCCT-----GAGAAAGCCAGAGGAAGCATCCAAA	3861
Oy	1266	GlyHlaValIleTyrGlnGlyLeuGlyGlyHlaIleLeuSerTyrGlnGlyGlyMetSer	1285
Db	3862	GGCCATGATTATTGTAAAGGCAAAAGTGCAATATCTTGTCATATGATAATTAAGAAAT	3921
Oy	1286	ValThrGlnCysSerLeuGluAspGlyAArgSerSerGlyProProHlaIleGluThrAla	1305
Db	3922	GCC-----CGAAGAGGAGCTAGAGTCCAGACACAGCTCAAGAAATCAGT	3966
Oy	1306	AlaProLysArgThrTyrAspMetMetGlnGlyArgValGlyArgAlaIleSer-----	1323
Db	3967	TTA---AAGAGACCTATGATTCAGTGAAGAAATTAATTAAGCAAGGATCTCATAGAG	4023
Oy	1324	-----SerAlaSerIleGlnGlyLeuMetGlyArgAlaIleProProGluArg	1339
Db	4024	GAGTCTCCTGTATCAGCACCGTTAGAGGGCTGATATGCGCAGCATTAACC-----AGG	4077
Oy	1340	HlaSerProHla--HlaLeuGlyGlnGlnHlaIleAlaGlySerIleThrGlnGly	1358
Db	4078	GGAGATCTCATTTGACCTCAAAAGAAAGAGCTGATTTGCTGGCTCCATATACAGAGGG	4137
Oy	1359	IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLeuLeuLeu	1378

Db	4138	ACACCAAGACAAACAATGAAAGCTTTGAAATGGCTTT---AAATATCCAAACAAATT	4194
Oy	1379	LYSARGGLUGLYThrProProProProProSerArgAspLeuThrGluAlaTyrLys	1398
Db	4195	AAAAGGAAAGT-----CTCCCATACAGCATTTGAAGTGCCATT---	4233
Oy	1399	ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal	1418
Db	4237	-----ACCAAGGAAGAACCA---TATGATGGC---ATCCACCACTTC	4272
Oy	1419	LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeu-----	1433
Db	4273	AAAGAAATGGGGCGTTTCATTATCATAGATGTTCCAAAGGCACAAATATTTAACTCAGAAAGT	4333
Oy	1434	ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln	1455
Db	4333	CGAAAAATCTCCAGAAAGTGCTCAGAGCACACGGCGGAATATGAGGGTTCATTCCAG	4399
Oy	1454	GlyThrProLeuLysTyrAspThrGlyValAserThrThrGlySerLysHisAspVal	1477
Db	4393	GGCACACCAATAAAAGTTTGAACAACAC---TCAGGTCAATGGCATAAACAAATGTC	4445
Oy	1474	ArgSerLeuIleGlySerProGlyValArgThrPheProValHisProLeuAspValMet	1493
Db	4450	AAATCTTAAATCACGGGGCTTACGAAACTATCCCGTGAATAGCTCCGCTGAATTTGTG	4509
Oy	1494	AlaAsp---AlaArgAlaLeuGluArgAlaLeuTyrGlu-----GlnSer	1507
Db	4510	CCAGAGACATAAAGAGTGTGAACCGGGAAATATGAGATGTGAAACAGCGGAGACC	4565
Oy	1508	LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro	1527
Db	4570	GTGGCTTCCCGGCACACGTCAAGTGTAAGCTTGCCCTCCCTCTTAAAGTCCACA---	4626
Oy	1528	ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly	1547
Db	4627	-----CTGCATGAAGCTCCAAAGACAACTGAGCCCTGGAGTTTATGATGACACCACT	4680
Oy	1548	Ala-----ProPheAlaGlyHisIleuProAlaArgGlySerProValThrMet	1563
Db	4681	GCACGGAGACCCCTGTGATGTTCAAAACCACTGCCAAGGCTCACCATATGAC	4740
Oy	1563	ArgGluProThrProArgLeuGluGlySerLeuSerSerSerLysAlaSerGlnAsp	1582
Db	4741	AGAACTTGTATGTTACAAATTCCT-----CTTAACAAGTCTACCAATCATGAA	4788
Oy	1583	ArgLysLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer	1598
Db	4789	AGGAATGTGACATGACCCCTTACCCAGAGGAAAGTATCCAGCGAAGTCTCCAGTGCT	4848
Oy	1599	ThrValProGluHisIleProHisProIleSerProTyrGluHisIleuLeuAlaGlyVal	1618
Db	4849	GGGGTGAACCTGTCTGAAGCCAC---AGTCCGTTTATCCCATCAACAGGGAGCC	4902
Oy	1619	SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro	1638
Db	4903	ACTCGAGCGAGGTTTATTGAGGCCACCTGCCCCAGCAATTGATTCAC---GCCATGCT	4959
Oy	1639	ArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArgHisIleuAlaPro	1657
Db	4960	TTTCAACAGGGCTTTGATCTCGACGCGGCTTACTGTTTCAAGACAGCTTTCACCA	5019
Oy	1658	AsnProThrTyrProHisIleuTyrProProTyrIleuIleArgGlyTyrProAspThrAla	1677
Db	5020	ACTTCAGATTAAACCAATCAGTATCAAGCTTTAC-----	5052
Oy	1678	AlaLeuGluAsn---ArgGlnThrIleIleAspAspTyrIleThrSerGlnMetHis	1696
Db	5053	GCATGTGAAGAACACACACACATCTTAAATGATTAATTAACCTTACACACAGATGCAA	5112
Oy	1697	HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLysSerProArg	1716

Db 5113 GTGAAGCTTG-----CGTCCAGATGTGGCCAGAGCACTCTCCCAAGA 5154
Qy 1717 GlusSerSerLeuAlaLeuAnThrAlaAlaGlyProArgGlyLysLeuSerLeuSerGln 1736
Db 5155 GAGCAGACCCTGGTCTCCCATACCCAGCA---ACGAGAGAAATCTTAAGCTGACCAAT 5211
Qy 1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAap 1756
Db 5212 ATGCCTCCA---ACATTTTAAAGCTCTCATCCAGGGGGAACAAGCACTCTCCCATGAGC 5268
Qy 1757 ArgLeuAlaTyLeuProThrAlaProGlnProPhSerSerArg---HisSerSerSer 1775
Db 5269 AGAATCACTTAATATCTCGTACACAGATTACTTCTCCACAGCGGCACTCTGCT 5328
Qy 1776 ProLeuSerProGlyValProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 5329 TCCATGTCTCCAGACACCCACACACTT-----GCAGTCTCTGCAAGCTGTGAG 5379
Qy 1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlySerLysLeu 1815
Db 5380 AGGAAACGGGAACGGGAGCGGAGAGAGCGGAGACGGGAACG-----ATTGCT 5430
Qy 1816 ThrSerThrThrThyValGluHisAlaProLysProLysProGlyThrGluGlnSerSer 1835
Db 5431 GCAGCTTCTCCGACCTC-----TACCTGCGGCGAGCTCAGAAACG----- 5472
Qy 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 5473 -----CTGCGCGACCTGGCACTCGA 5496
Qy 1856 HisAlaHisGlnHisSerProLysSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTCGCTCCCTCCCTCCCT---TCAGTAAGAACTCAGAGACCAATGTTGCAACAGADA 5553
Qy 1875 ProSerValLeuHisAsnThrGlyMetLysGlyLysLeuThrAlaValGluProSerLys 1894
Db 5554 CCCAGGTGTTTCCAGAGAAACCAATGGAACCAAGTGAATCAACCTTGATCCCACTGCT 5613
Qy 1895 ProThrValLeuArgSerThrSerThrSerProVal-----ArgProAla 1910
Db 5614 CAGCTACGAATCATGCGCACTGCTGCGGGGCGCTTCAATAGCCAAAGCGCTCCAGCC 5673
Qy 1911 AlaThrProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyPro 1930
Db 5674 TCCCGTTACAAACACTGCTGGGATGCC---CTGGCTGCTGTTGGATGCTGCGAGCTGCT 5730
Qy 1931 ThrLeuMetGluProValLeuLeuProLys-----GluAlaProArgValAla 1946
Db 5731 GCAACCCAGATGATGTGTCCAAACAAAAGAGATGAAGCATGAAGCTGCCAGGTTAGAA 5790
Qy 1947 ArgProGluArgProArgAlaAsp----- 1954
Db 5791 GAAATTTTGAAGAGCAGGTCAAGCAAGTTAGTGAACAGACAGCAAGCTAAGCAAGAAACC 5850
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLys 1962
Db 5851 CTGAGAGTGAGAAAGATCTGTTCAAGTGTTAACACTTCTTCAAGCTTCCAAAGTGCG 5910
Qy 1963 ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProSerLysGlySer 1979
Db 5911 AAGCCC-----CAGCTCATTTCTTCAGTAGTTATTCTGAGGCTGGAGAA 5955
Qy 1980 GluProArgProLeuValProProValSerGlyHisAlaThrThrLeuAlaArgThrProAla 1999
Db 5956 GATTAAGGGCT-----CTTCCAAATCCAGATATGAGAGAAAGCTAAGACCAAGAGG 6009
Qy 2000 LysAsn-----LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSer 2017
Db 6010 AAGACTACCACTTACTGACAGTAACCTCATPAGACGTATCATCACCGGCAAAATGTGCTCG 6069
Qy 2018 AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerLysGlnGluLeuGlu 2037
Db 6070 GACAAAGATGCGAGGAAAGTGCTCTCAAAAGTTCAAGACTTCTT----- 6114

Qy 2038 LeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluProValSer 2057
Db 6115 ---AGTACCTTATCTTTCACAGATGATGAACACCTAAGAGATGATTAAGGTGATAAGT 6171
Qy 2058 ProValSerSerProSerLeuThrHisValArgGlyLeuProLysHisLeuGluGluLeu 2077
Db 6172 CTGCGACGCTCACTCGCGCACCCAGAGAACTGACAGACCTATCAGCCAGAGGTTGTT 6231
Qy 2078 AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db 6232 AAGCAAAATCAAGCGGAAATATCTTACCAACAATATGAAGACCA----- 6279
Qy 2098 GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTACATCATATGACCA-----CAGCAGGAATCACCATCT 6315
Qy 2118 ProLeuLeuGlnThrAlaProLys-----ValLysGlyHis 2129
Db 6316 CCCCAACACACAGTGGCCCTTTCACAGAGAGGAATGGGCAAGTGGCCAGGACC 6375
Qy 2130 GlnArgValValThrLeuAlaGlnHisLysSerGluValLysThrGlnAspTyThrArg 2149
Db 6376 CATCGCTGATCACTGCTGATCATCATCTGTCAAATATATCACAAAGATTTGCTAGA 6435
Qy 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTy 2164
Db 6436 AATCAAGTTTCTCGCAGACTCCCGACAG-----CTTCCTTCTTACATTCAG 6486
Qy 2165 SerPheProGlyAla-----SerCysProValLeuAspLeuArgArgProSerAsp 2182
Db 6487 AACTCACTTCTGCTTGGTATCTACACCTG-----AGAGCTAAACATCAAC 6537
Qy 2183 LeuTyLeuProProProAsp-----HisGlyAlaPro---AlaArgGly 2196
Db 6538 CGTTACAGCCCAAGATCCAGGCTCAGTGTTCATCATCAACAAAGACAGGTTCAAGGCTC 6597
Qy 2197 SerProHis-----SerGluGlyLysArgSerProGluProAsnLys 2211
Db 6598 TCTCCGAAATCTTGTGACAAATCCAGGGGAAGAGCGCTGAAATCCCGCAGAGAG 6657
Qy 2212 ThrSerValLeuGlyGlyGlyLysArgLysLeuProValSerProProGluGlyMet 2231
Db 6658 AGTCACGTC-----TCTTCCAGCCCTACAGACCCATCTCCACCCAG----- 6702
Qy 2232 ThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGln 2251
Db 6703 GTTCCGTTGTGATGAAACAGACAGACAGCTTGTCTTGTCTCAGAGGGCGAGAG 6762
Qy 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPhe 2271
Db 6763 CTGCAAGACAGAGAAATGATCCCGCTCACAGGAGATGAAGTAACTTGGCTTCATTC 6822
Qy 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLeuAsn 2291
Db 6823 TTCACCAACCTT---GAAATACATCACCCATGTTAAATCAAAAGAAAGAGATTTT 6879
Qy 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyArgHisLysSerGlnProGlyThr 2311
Db 6880 CTAAGTTAACTCTCTGCTGAGGTGACTGTGATAGGACAGCTCTTACAGGAACT 6939
Qy 2312 GluLeuPheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAla 2331
Db 6940 GAGATCTTAATCTGACAGATTAATCATGACAGGCTCAGTTAGCTCTAGAGCCATTCT 6999
Qy 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleLeuArgLysAlaLeuMet 2351
Db 7000 TTGCTGATCTGCGCAGT---AATCTTGGCTGGAAGACATTAACGAAAGCTTCATG 7056
Qy 2352 GlyLysTyArgProLysProGluGlu-----SerProProLeuSerAla 2365
Db 7057 GAAAGCTTATGACAAAGTAAAGATCATGGAAGTTGATGATGCCAGCTTATGGAGATA 7116

QY	2366	AsnAlaPheAsnProLeuAsnAlaSerLeuAlaSerLeuProAlaAlaMetProIlePheThr	2365
Db	7117	-----GTGCTGTACTGTCACACCTTAGTT-----	-----GTGACCAAGT 7152
QY	2386	AlaAspGlyAArgSerAspHisThrLeuThrSerProGlyGlyGly-----	-----LysAla 2403
Db	7153	GGTAGACACCGAAGAGAGAGAGAGAGGCCCATCCTCACTTACAGAGAGATTTCGAACA	7212
QY	2404	LysValSerGlyAArgProSerSerArgLysAlaLysSerProAlaPro-----	-----GlyLeu 2422
Db	7213	AAGCTGATCAGCAAGCTCAACACGACGAGAAATCTAAGTCTCCTATGCTGGCAAGGCTAC	7272
QY	2442	AlaSerGlyAAspArgProProSerValSerSerValHisSerGlyGlyAAspCysAsnArg	2441
Db	7273	TTAGGAACGGAACGGCCCTCTTCACATCTCTCTGTACATTCAGAAAGGGGATTAACATAGG	7332
QY	2442	ArgThrProLeuThrAsnArgValTrpGlyAspArgProSerSerAlaGlySerThrPro	2461
Db	7333	CAGACGCCA-----GGGTGGGCGCTGGGAAGACAGGCCCTCTTCAACAGGCTCAACTAG	7366
QY	2462	PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPro	2481
Db	7387	TTTCTGTATTAACCCCTGACTGATCGGATG-----CTCAGCAGATCTCCACCAAC	7437
QY	2482	ProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----	-----HisAla 2497
Db	7438	CCGATTGCATGTCTCCCTCTGGCGGTGAACCAAGACGCTCTCTCAACACAGACAGATC	7497
QY	2498	TrpAspGlnGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu	2517
Db	7498	TGGAGCGACAGACGCTGCCCACTGCTCTCACACACATACGAGACCTCTCGGATATGTGAT	7557
RESULT 13			
AX578069			
LOCUS	AX578069	7940 bp	DNA
DEFINITION	Sequence 191 from Patent WO02081745.		linear
ACCESSION	AX578069		
VERSION	AX578069.1	GI:27647277	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	1 Garcia,T., roman Roman,S., Baron,R., Call,K., Theilhaber,J.,		
TITLE	Connolly,T., Jackson,A., Bushnell,S.E. and Rawadi,G.		
JOURNAL	Genes involved in osteogenesis, and methods of use		
FEATURES	Patent: WO 02081745-A 191 17-OCT-2002;		
	Aventis Pharma S.A. (FR)		
	Location/Qualifiers		
source	1..7940		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	/note="Homo sapiens nuclear receptor co-repressor 1		
	(NCOR1), mRNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.07e-75	length:	7940
Score:	4147.50	Matches:	1099
Percent Similarity:	53.33%	Conservative:	341
Best Local Similarity:	40.70%	Mismatches:	785
Query Match:	31.38%	Indels:	475
Db:	6	Gaps:	108
US-09-522-753-5 (1-2517) x AX578069 (1-7940)			
QY	16	GluProAlaGlyTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr	35
Db	289	CAAGGTCTGTATCTCTCCCTCACTGTCTCAGATATCATTTCCCAACACCGCCACACAG	348
QY	36	AspValGlyLeuLeuGluTyrGln-----HisHisSerAlaGArgTyrAlaSerHisLeu	53

Db 349 GAGTTCGAGTCCCTGATTATTCCTCTATCTTGAAGTACAGGCACTACAGCTT 408
OY 54 SerProGlySerIleIleGlnProGlnArgArgProSerIleuSerGluPheGln 73
Db 409 TTGCAGCAACAGCAGCAGCAGCAGCTTGGAAGGCGACTCTTGCTTCAGAAATTTAC 468
OY 74 ProGlyIleGlnArgSerGlnGluLeuHISleuArgProGluSerHisSerTyrIleuPro 93
Db 469 CCAGGTTCTTCAGAGGCTTCAGAA-----AGGAAACTAGTTTGAACCTTTCACTCA 522
OY 94 GluIleuGlyIleSerGluMetGluPheIleGluSerIleValArgProArgIleuGluLeu 113
Db 523 GAGCCATCCCAAGTGCATGATTCATCTGAAATCGAAGCACAACGCTTGGAACAGTT 582
OY 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 583 TCTGATTCATATTTTCAGCGTGCAGTGCCTGGTTTTGCTTTAGTGCACCGGCTGCA 642
OY 123 LeuLeuAlaIleArgIleGlnProAlaGlySerGluAspleuThrIleValArgSerIleuThr 142
Db 643 -----GAAGGGCTGAGGGCT---TCTGCATGCTGAAGAGATCCAGCATTCGCA 690
OY 143 GlyIleuSerGlu---ProValSerProSerProProHisIleThrArgProGluLeuGlu 161
Db 691 GGCAAACTGTGAAGCTCCATCTCTTCCAATTTGGGGCAACCATGTGCAGATGATCAAAAT 750
OY 162 LeuValProProArgLeuSerIleGluIleuIleGlnHisMetAspArgValAspArg 181
Db 751 GCTTCACCTTCAAAACTCTCAAGAAAGATTAATACAGAGTATGAGTTCGTGTAGTCA 810
OY 182 GluIleIleIleMetValGluGlnGlnIleSerIleValIleuValIleGlnGlnIleuGlu 201
Db 811 GAAATTCGAAAGATGAGAACAGCAATCTTAACTGAAAAAGAAACAACAACACTTCGA 870
OY 202 GluGluIleAlaIleValProProGluIleProGluIleValSerProProProIleGluSer 221
Db 871 GAAAGGAGGCTAAACCTCTGACCTGAGAACCCGCTGCTCTCTCTCTGTTGAGCAG 930
OY 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValIleGluAlaAla 241
Db 931 AAACACGCGAGTATGTGTCCAAATTTATATGATGAGAAATGGAAAAAGCAAGAAAGCT 990
OY 242 HisArgIleIleuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 991 CATAAAAATTTTGAAGGCTTGGGCCCAAAAGTTGACCTGCACGTATTAACCAACACTCA 1050
OY 262 AspIleIleArgIleIleGluAsnIleIleValIleAsnGlnIleMetArgIleValIleIle 281
Db 1051 GATACCAAGGTGACCATGAGAACATCAAGCAAAACCAAGGTGATGAGAGAAAAACTCATTT 1110
OY 282 LeuTyrPheIleValArgIleAsnHisAlaArgIleGlnIleIleTyrIleValIlePheCysGlnArg 301
Db 1111 TTATTTTAAATAAATAAATATCATGCAAGAAACAAAGGAAACAAAAATCTGCACGCT 1170
OY 302 TyrAspGluIleuMetGluAlaLeuGlnIleValIleValIleGluIleGluAsnAsnProArg 321
Db 1171 TATGATCAGCTCATGAGAGCGCATGGGAAAAAAAGTGCACAGAAATGAATAATCTCCGG 1230
OY 322 ArgArgIleAlaValSerIleValArgGluTyrTyrGluIleuSerGlnPheProGluIleArg 341
Db 1231 AGGAAAGCTAAAGAAACAAAAACAAGGAAATCATGTAAAAAGCATTTCCAGAAATTCGA 1290
OY 342 LysGlnArgIleuLeuGlnIleArgMetGlnSerArgValGlyIleArgIleSerGlyLeu 361
Db 1291 AAACAAAGAGAACAGCAAGAAAGATTTCAG---GAGTTGGCGAGAGGGGAGCTGTCTT 1347
OY 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
Db 1348 TCAGCCCAACCATTCGTAGAGATGAGCAAGAAATTTCTGAATTTATGTATGGAGCTTCGAG 1407
OY 382 GlnGluAsnIleuGlnIleValIleMetArgGlnIleuAlaValIleProProMetLeuTyrAsp 401

Db 1408 CAGGAGATAATATGAGAAACAATGCGGACGCTCTGTGATTCACCTATGATGTTGAT 1467
Qy 402 AIAAspGlnGlnArgIleIysPheIleAsnMetAsnGlyLeuMetAlaAspPromeIys 421
Db 1468 GCAGAAACAAGACGATCAAGTTCATTAACTGAATGGCTTATGAGAGACCTATAGAA 1527
Qy 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnIleuIysGluThrPheArg 441
Db 1528 GTGTATAAAGATAGGACGTTTATGTAATGTTTGACATGACCATGAAAAAGAGATCTTAAAG 1587
Qy 442 GlnIysPheMetGlnHisProIysAsnPheGlyLeuIleAlaSerPheIleuIysGly 461
Db 1588 GACAACTTATTCAGATCCAAAAAATTGACTATTTGACATATGCACTTGGAGAGAAAG 1647
Qy 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysAsnGlnAsnTyrLysSer 481
Db 1648 AGGTTCCTCGATTGCTTTTGTATTTACTATTTAACCAAGAAAAATAGAAATTTAAAGCC 1707
Qy 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnIleuIleuIleuIleu 500
Db 1708 CTCTGACAGAAAGATATGAGAAACGACAGGACGAAACGACAAATTGCTCGACCTCG 1767
Qy 501 Gln 520
Db 1768 CAGAGAGAAAGATAGAGAGAAAGAGAG---GATTAAGCAGAAAAAACAGAAAAA 1821
Qy 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValGluAsn 540
Db 1822 AAAGAAAGAAAGAAAGAAAGATGAAAGAGAAAGAAAGATGAAAGAAAGAAAGAAAGAAAT 1881
Qy 541 AspLysGluAspLeuLeuLysGluLysTyrThrAspAspThrSerGlyLysPheAsnAspGlu 560
Db 1882 ACCAAGGAA---AAGGACAAAGATAGATGATACGACG---GAGAAATCTGAGGAA 1929
Qy 561 LysGluAlaValAlaSerLysGlyArgGlyLysThrAlaAsnSerGlnGlnIysArgGlyLys 580
Db 1930 AGAGACCAAGCCACACCCGCGGGGCGAAAGACTGCCAAGCTGAGGCGCGCTTAAGAGG 1989
Qy 581 ArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnGlnAlaIleThrProGlnGln 600
Db 1990 CGGATCACCAAGCTCATGACAAAGCAAGCTGCAGCTGCCAGCTCAGCGGACGGCGCT 2049
Qy 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
Db 2050 ACTGAAAGACCCCAACCACTCTGCCACCGCCACCAAGAAACCACTTCTACAGAGCTGTG 2109
Qy 611 GlnSerSerArgTrpThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 630
Db 2110 GAGACCTCTCGATGACAGAGAGAGAAATGGAGCTGTAAAGAAAGCTTAAGTAAGACAT 2169
Qy 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 2170 GGTCGTAACCTGGGACCAACTTGTCTAAATGTGTGGAAACGAAAGTGAAGCTCAATTTAA 2229
Qy 651 AsnPheTyrPheAsnTyrIleLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 670
Db 2230 AACTTCTATTATTTAACTATAAAGCGACACAATCTTGACCAACCTCTTACAGCGACATAAA 2289
Qy 671 LeuLysMetGluLysGlnArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAla 690
Db 2290 CAGAAAACTTCACGAAAACTCTGTGAAGCGAGATGTCTCTAAGTGAAGATGTCTGCT 2349
Qy 691 SerGlnGlnAlaAlaPheProProValValGluAspGlnGlnIleuMetGlnAlaSerGlyVal 710
Db 2350 TCACACTGTTTCTGCT-----CAGAGAGATGAAGATTTGAAGCTCC----- 2391
Qy 711 SerGlyAsnGlnGlnIleuMetValGlnGlnAlaGlnAlaLeuHisAlaSerGlyAsnGln 730
Db 2392 ---AATGAAGAAAGAAATCCAGAAAGACAGCGAA----- 2421
Qy 731 ValProArgGlyLysLysSerGlyProAlaThrValAsnAsnSerSerAspThrGlnSer 750
Db 2421 ----- 2421

Qy 751 IleProSerProHisThrGlnAlaIleLysAspThrGlnIleAsnGlyProLysProPro 770
Db 2422 -----GTGGAAGCTGTCAAG----- 2436
Qy 771 AlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrSer 790
Db 2437 -----CCACGAGAGACAGTCTGAAAAATGCTACTTCT 2469
Qy 791 ArgAlaProIleGluProThrProAlaSerGlnAlaThrGlyAlaProThrProProPro 810
Db 2470 CAGAGAAACAGAAACCTCGGTTGAGCTTGAAG-----CCAGCCAGCGAAACT 2517
Qy 811 AlaPro---ProSerProSer---AlaProProProValValProLysGlnGlnLysGlu 828
Db 2518 GCACCCAGTACATCTCCCTCTTACAGATTCCAAGTACAAACAGCTGAAGATGAAGCT 2577
Qy 829 GlnGluThrAlaAlaAlaProProVal----- 837
Db 2578 GTGAGAACCCAGGTGAATGACAGCATCACTGCTGAGACAGCAGACAGATGATGAT 2637
Qy 838 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 851
Db 2638 CAGCAGAGACAGATGCTGAAGAGGTTCTGTTGTGATGCCCCACCGCTACCAAA--- 2694
Qy 852 AlaValAspThrGlyLysAlaGlnGluProValLysSerGlnCysThrGlnGlnAlaGlu 871
Db 2695 ---GCTGACTCTGTGACCTGTAAGTGAAGGTCGACAGAAACCATGATCTAAAGTTGAA 2751
Qy 872 GlnGlyProAlaLysGlyLysAspAlaGlnAlaIleGlnAlaThrAlaGlnGlyAlaLeu 891
Db 2752 GGTGATATATCAAGAAAGAAAGACTTGATGATAGCC----- 2787
Qy 892 LysAlaGlnLysLysGlnGlyLysSerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
Db 2788 ---AGTGAAGAGTGAACCTAGATGAAGATTTGTGTGTATGCTCAGCAAAATTAATGCC 2844
Qy 912 -----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
Db 2845 CAAAGCCCGAGCCCGACAGACAGACATGATTCAGTGCCAGCTGACGCGCTGATGAG--- 2901
Qy 927 AspGlnAlaGlnGlyLysAspLysAsnArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGAGAGCCAGAGAGGCGAGAGATGTTTCTATGAGACTCAAAAGCTTCA 2958
Qy 944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu 961
Db 2959 CTGTTAAACCCCACTGGATCTATCTGCTC---TCATCTCGTTTAAACCAATTCACATG 3015
Qy 962 AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle----- 976
Db 3016 GATCTGCCACAGACTTCAGCATGAGCTGCTGTATTCACCAATGATGTATCTGCACCCCA 3075
Qy 977 -----GlnValThr 979
Db 3076 TGTAAACATCAATTTGAACCCCAAGTGAAGCGGCTATCTCTTACAGGACACATTTAA 3135
Qy 980 LysValHisGluProProArgGlnAspAlaAlaProThrLysProAlaProProAlaPro 999
Db 3136 GCAATGACATGATCAAGCACTCTGTGAG----- 3162
Qy 1000 ProProProGlnAsnLeuGlnProGlnUserAspAlaProGlnGlnProGlySerSerPro 1019
Db 3163 ---GAGCAGCGGACAGACAAAGAAAGATGTTGGATGTAAGATTTCTCAAGTCCA 3219
Qy 1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGlnAlaPheAlaIleGlnAla 1039
Db 3220 TGTGGCATTCCAAGAGTCCA-----AACAGAGAG----- 3249
Qy 1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 3250 -----TGG----- 3252

QY	1060	ArgGluVal11LeuValSerProHisValProAerProSerAlaPheSerTyrAlaPro	1079	1099
Db	3253	---GAAGTCTTCCAGCTCCAGCTCCACAT-----CAATTGATTAATAATCTCCCT	3297	
QY	1080	ProGluHisProLeuProLeuGluYLeuHisValPheTyrAlaArgProValLeuProArgPro	1099	
Db	3288	GAAGGCGGTTGGGCTTCCG-----ACAACTCGACCAACGAGCCCGCC	3342	
QY	1100	ProThrHisSerPheProProProLeuHisLeuSerAlaYLeuHisProSerValLeuGlu	1119	
Db	3343	CGTTCATCCCGGTATCCAAACCAACAGAGGCTTCAAAAA---CAATCTTTTATA---	3396	
QY	1120	ArgGluHisGlyAlaHisSerGluGly-----MetSerValGluLeuHis	1134	
Db	3397	---ATGGAGGGCTTCATCTATCCACAGGAACACCGACGCTTAATTGACCTTCATATACAG	3453	
QY	1135	ValProTyrSerGluHisAlaValAlaPro---ValGlyProValThiMetGlyLeuPro	1153	
Db	3454	CGTTCCTTAACSTCAAGAAACACCCGACGGCTGAGTAAAGATATATCTCTTGAGACTGCA	3513	
QY	1154	LeuProMetAerProYLeuValLeuValaProPhePheSerGlyValYLeuGluGluMetSer	1173	
Db	3514	CGGCAACAGGAATCTGCCAAATCAGATCTTTGGCTCATACAGACAGAAATTTCT	3573	
QY	1174	ProArgGlyGluAlaGlyProProGluSerLeuGlyValProThrAlaGluAlaSer	1193	
Db	3574	CCCCGAAACCAAACTCAACCTGAGGGGTCTGTGGTC---AGGGCCCAATGAAAGT	3630	
QY	1194	ValLeuValArgGlyThrAlaLeuGlySerValProGlyGlySerHisThrGlyGlyLePro	1213	
Db	3631	GTAAGCAGAGTACCGCA---GAGACCATACAAAGAAAGATTAATCTCGGGAACTCCA	3687	
QY	1214	SerThrArgValProSerAerSerAlaHisThrTyrArgGlySerHisThrHisGlyThr	1233	
Db	3688	ACCGACCAAAATTTCACTGAGAGACATTCCTCATCCCGGGCTTATATCTGAGGCAACC	3747	
QY	1234	ProAla-----AerValLeuTyrGlyGlyThrHisThrArg	1245	
Db	3748	CGGGTCTGCCCGACATCGGCATACCAACAGAGCTTTGGTGAAGGGTTCATTTCCGAGA	3807	
QY	1246	HisLeuGlyLeuAerSerProSerArgLeuAerArgGlyValArgGlyAerProLeuY	1265	
Db	3808	ATGCCCATGGAAGACAGCATGCT---GAGAAAGCAAGAGAAAGCTGCATCCAA	3861	
QY	1266	GluHisValHisTyrGluGlyGlyLeuValLeuHisValLeuSerTyrGluGlyGlyMetSer	1285	
Db	3862	GCGCATGTTATTTAAGGCAAAAGTGAACATATCTTGATATGATTAATTAAGAT	3921	
QY	1286	ValThrGluYerSerLeuAerProYArgSerSerSerGlyProProHisGlyThrAla	1305	
Db	3922	GCC-----CGAAGAGGGACTAGAGTCCAAACAGCTCAAGTAATACGT	3966	
QY	1306	AlaProValArgThrTyrAerMetMetGluGlyValGlyAlaGlyAlaSer-----	1323	
Db	3967	TTA---AAGAGAGCTATGAATCATGTGAGAAATATAAAGCAAGGATGTCAATGAGG	4023	
QY	1324	-----SerAlaSerHisGluGlyLeuMetGlyValGlyAlaHisProProGluValArg	1339	
Db	4024	GAGTCTCTGTATACGACCGTGTAGAGGGCTGATATGCCAGATTAATCC---AGG	4077	
QY	1340	HisSerProHis---HisLeuValGluGluHisHisHisHisLeuArgGlySerHisThrGluGly	1358	
Db	4078	GGAGTCTCTATTTGCACTCAAAAGAAAGCACTGATCTGTGGCTCCATATATGCAAGGG	4137	
QY	1359	HisProArgSerTyrValGluAlaGluGluAerTyrLeuAerArgGlyAlaValLeuLeu	1378	
Db	4138	ACACCAAGACCAACAACTGAAGAACTTTGATGATGGCTT---AAATATCCCAACAAATT	4194	
QY	1379	LeuAerArgGlyGlyThrProProProProProSerAerAerAerLeuThrGluAlaTyrLeu	1398	
Db	4195	AAAGGGGAAAGT-----CTCTCCCATGACGATTTGAAGGTGCCATT--	4236	
QY	1399	ThrGluAlaLeuGlyProLeuValLeuValAerProAlaHisGluGlyLeuValAlaThrVal	1418	

Db	4237	-----ACCAAGGAAACCA---TATGATGC---ATCACCACATC	1433	4272
Oy	1419	LYGGLIAGLYARGSERILEISGLILEPROARGGLUGLULE-----	1433	4273
Db	4273	AAAGAAATGGGGCTTCATTCATGAGATTCGAAAGCAAGATATTTAACTCAGAAAGT	1433	4273
Oy	1434	ARGHLETHRPROGLULEPROLEUALBROARGPROLEULVGLIYSEILETHGLIN	1433	4273
Db	4333	CGAAATACTCAGAAAGTGCTCCAGAGCAACGGCCGATTAATTGAGGGTTTCATTTCCAG	1433	4273
Oy	1454	GLYTHRPROLEULVETRYRLEPETHGLYALASERTHRTGGLYSELYVLYNHIAEPYAL	1433	4273
Db	4393	GGCACACCAATAAGTTTGACCAACAC--TCAGGTCAATGTCATTCACCAATATGTC	1433	4273
Oy	1474	ARGSERLEULIEGLYSERPROGLYARGTHRHEPROVALHLEPROLEUALBROVALMET	1433	4273
Db	4450	AAATCTCTTAATACAGGGGCTTACGAAATATTCCTGGATATGCTCCGCTGAAATGTG	1433	4273
Oy	1494	ALAAEP--ALARGALALEUGLUALARGALCYETRYGLU-----GLUSER	1433	4273
Db	4510	CCAGGAACATAAAGTGTTGAAAGGGGAAATATGAGATGTGAAAGCAGGGAGAGCC	1433	4273
Oy	1508	LEULYSERARGPROGLYTHRALSERSESERGLYSEILEALARGLYALBRO	1433	4273
Db	4570	GTGGCTTCCGGCACACGTGAGTGAAGCTCTGGCCCTCGTTTGAAGTCCACA--	1433	4273
Oy	1528	VALLEVALPROGLULEUGLYLYPROARGHLEPROLEUTHRYTRGLUALBROHISGLY	1433	4273
Db	4627	-----CTGCATGAAGCTCCCAAGCACACATGAGCCCTGGATTTTGAATGACACCAAT	1433	4273
Oy	1548	ALA-----PROPHALAGLYHIALEUPROARGGLYSEPROVALTHMET	1433	4273
Db	4661	GCACGGAGGAGCCCTGTGAGTATTCACAAACCAATGTCCAGAGGCTACCCATGATGAC	1433	4273
Oy	1563	ARGGLUPROTHRPROARGLEUGLUGLYSERLEUSERSERLYVALSERGLINBP	1433	4273
Db	4741	AGAACTTCTGATGTACATATTCCT-----CTTAACAAGTCTAACATATGAA	1433	4273
Oy	1583	ARGLYLEUTHRSERTHPRORARGLU-----ILEALYSEPROHISER	1433	4273
Db	4789	AGGAAATCGACATGACCCCTTACCAAGGAAAGTATCCAGGAAAGTCTCCAGTCT	1433	4273
Oy	1599	THRYVALPROGLUHIHLEPROHIALEPROLSEERPROTYRGLUHIHLEULVARGLYVAL	1433	4273
Db	4849	GGGATGACCTCGTCGTGAGCCAC-----AGTCCCTTTATCCCATCACAGAGCAGC	1433	4273
Oy	1619	SERGLYVALBPYLEUTHRYARGSERHISILEPROLEUALBROHLEBPROTHRSEILEPRO	1433	4273
Db	4903	ACTCGAGGCGAGGTTTATTTGAGGCACCTGCCACGCAATGATGCA---GCCATGCT	1433	4273
Oy	1639	ARGGLYLEPROLEUALBP--ALALALALALARYTRYLEUPROARGHISLEUALBRO	1433	4273
Db	4960	TTTCACAGGGCTTTGATCTCTGACGGCGCTTAACTGTTTCAGAAACAGTTTCAACA	1433	4273
Oy	1658	ASNPROTHRTRYPROHISLEUTHRYPROBROTYRLEULBARGGLYTRYPROABPTHRALA	1433	4273
Db	5020	ACTCCAGGTTTACCAAGTCAAGTATCACTTAC-----	1433	4273
Oy	1678	ALALEUGLUN--ARGLINTHRIELEASNAAPTRYLLETHRSERGLINMETHIS	1433	4273
Db	5053	GCATGTGGAACAACAAGACAGCAATTTAAATGATTAATCACTTACCTACAAACAGATGCA	1433	4273
Oy	1697	HIAENHTRALATHRALAMEVALAGLARGALABPMELEUALBROGLYLEUSERPROARG	1433	4273
Db	5113	GTGAACTTG-----CGTCCAGATGTGCAAGAGAGCTCTCCCAAGA	1433	4273
Oy	1717	GLUSERSERLEUALALEUAANTYTRALALAGLYPROARGGLYLEILEAPLEUSERGLIN	1433	4273
Db	5155	GAGACGACCTGGGTCTCCATACCCAGCA---ACGAGAGAAATCATTTGACCTGACCAAT	1433	4273
Oy	1737	VALPROHISLEUPROVALLEUALPROPROTHRPROGLYTHRPROALATHRALAMEALBP	1433	4273

[illegible]

QY	2078	AspIysSerH1sleuGIuGIygluLeuAArgProIySeGIuProGIyProValysleuGIy	2077
Db	6232	AAGCAAAATCAAGCGAATAATGATCTTACAGAGAAATATGAAGGACA-----	6279
QY	2098	GIyGIuIaIaIaH1sleuProH1sleuAArgProleuProGIuSerGIuProSerSer	2117
Db	6280	-----TTACATCACTATTCAGCA-----CAGCAGGAATCACCATCT	6315
QY	2118	ProleuGIuInthraIaProGIy-----ValysGIyH1s	2129
Db	6316	CCCCAACAAACGTCGCCCCCTTCTTTCACAGGCAAGGAAATGGGGCAAGTCCAGAC	6375
QY	2130	GIuHArgValIaIThIreIaIaGIuH1sIleSerGIuValIeThIrgIaAspIyThArg	2149
Db	6376	CATGGCGTGAATCACTTCGTGATCAATCTGTCAAAATTATCACAACAAGATTTTGCTGA	6435
QY	2150	H1sH1s-----ProGIuGIuLeuSerIaProleuProIaIaProleuTy	2164
Db	6436	AATCAAGTTCTCTGCAGACTCCCCAGAG-----CCTCTACTCTTCATCTCCAG	6486
QY	2165	SerIheProGIyAla-----SerCyProValIeAspIeAspIaArgProSerAsp	2182
Db	6487	AACTCACCTTCGTGGTTGGTATTCACACTGTG-----AGGACTAAACATCAAC	6537
QY	2183	LeuTyIleuProProProAsp-----H1sGIyIaPro--AlaArgIy	2196
Db	6538	CGTTACAGCCCAAGATCCACAGCTCAGTCTTCATCATCAAAACACAGGTTCAAGGCT	6597
QY	2197	SerProH1s-----SerGIuGIyIyAsArgSerProGIuProAsnIys	2211
Db	6598	TCTCCAGAAATCTTGTGACAAATCCAGGGGAAGTGGCCCTGGAAATCCCCAGAGAG	6657
QY	2212	ThIseValIeugIyGIyGIaAspIyIleGIuProValSerProProGIuGIyMet	2231
Db	6658	AGTCACGTC-----TCTCCGAGCCCTACAGACCCCATCTCCACCCCA-----	6702
QY	2232	ThrGIuProGIyH1sSerArgSerAlaValTyProleuLeuTyArgAspGIyGIuIn	2251
Db	6703	GTTCCGGTTGCGATGAGAAACAGACAGCTTGCCTGTTGTCTACAGGGGCGCAGAG	6762
QY	2252	ThrGIuProSerAArgMeGIySerIySerProGIyAsnThIrgIaInProProAlaPhe	2271
Db	6763	CCTGCAGACGACGAATGATGCCCGCTCACAGAGAGATATACCTACTTCCCTTCATTC	6822
QY	2272	PheSerIyIeLeuThIrgIuSerAsnSerAlaMetValIySerIyIyGIuInIleAsn	2291
Db	6823	TTTCAACAAGCTT---GAAATATACATCACCCATGTTAATCAAGGAACAGAGATTATT	6879
QY	2292	IyIeIyIeLeuAnThrH1sAsnArgAsnGIuProGIuTyArgenIleSerGIuProGIyTh	2311
Db	6880	CGTAAGTTGAACCTCTGTGTGAGAGTCACTCTATATGGCAGCTGTCCACCCGAACT	6939
QY	2312	GIuIePheAsnMetProAlaIleThIeThGIyGIuMetThTyArgSerGIuAla	2331
Db	6940	GAGATCTTTATCTGCACGACGATTAACGTACAGGCTCAGTTAGCTAGAGCCATTCT	6999
QY	2332	ValGIuGIuH1sAlaSerThAsnMetGIyLeuGIuIaIeIleArgIyAlaLeuMet	2351
Db	7000	TTTGGCTATCTCCGCAGT---AATCTTGGGTGGAAGCATTTATCAGGAAGCTCTCATG	7056
QY	2352	GIyIySerTyArgArgIuInTPGIuGIu-----SerProProLeuSerAla	2365
Db	7057	GGAAGCTTGTATGACAAAGTTGAGATCAGGAGTTGTCAATGCCAGCCATAGGAGTA	7116
QY	2366	AsnAlaPheAsnProLeuAsnAlaSerAlaSerIeProAlaIaIaMetProIleThraIa	2385
Db	7117	-----GTGGCTGTGATCGCAACACCTCAGTT-----GTGACCACT	7152
QY	2386	AlaAspGIyArgSerAspH1sThIeThIserProGIyGIyGIyGIy-----IyAla	2403
Db	7153	GGTGGACCAACAAGAGAAAGGGAGCCCATCACTCATTCAGAGAGGATTTGCAACAACA	7212


```

QY 162 LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg 181
DB 751 GGTTCACCTTCAAAATCTCAAAAGAGATTAATACAGATATGATGCTGTAGATCA 810
QY 182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGlu 201
DB 811 GAAATTCGCAAAAGTAAACAGCAGATCTTAAACTCAAAAGAAACAAACAGCTGGAA 870
QY 202 GluGluAlaAlaLysProGluProGluLysProValSerProProIleGluSer 221
DB 871 GAAAGGCGCTTAACCTCTGAGCTGAGAGCCCGTGTCTCTCTCTCTGAGCGCAG 930
QY 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaAla 241
DB 931 AAACACCGCAGATATGTCGCAAAATTAATTATGATGACATCGGAAAAGAGAAAGCT 990
QY 242 HisArgIleLeuGluGluLysGluProGlnValGluLeuProLeuTyrAsnGlnProSer 261
DB 991 CATAAATTTTGAAGCTCTTGCCCAAAAGTTGAACCTGCTATTAACCGCATCA 1050
QY 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
DB 1051 GATACCAAGGTGATCATGAGAACATCAAGACAAACGAGTGAAGAAAACCTCAT 1110
QY 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArg 301
DB 1111 TTATTTTAAAGAAAGAAATCATGCAAGAAAACAAAGGGAACAAATAATCTGCCAGCT 1170
QY 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
DB 1171 TATGATCAGCTCTGAGGAGCATGCGAGAAAAGTGAGACGAATTAATAATCTCTCG 1230
QY 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
DB 1231 AGGAAAGCTTAAGAAAGCAAAACAGAGAACTATGAAAGCAAGTTTCCAGAAATTCAG 1290
QY 342 LysGlnArgGluLeuGlnLysGlnMetGlnSerArgValGlyGlnArgLysSerGlyLeu 361
DB 1291 AAACAAAGAGAACAGCAAGAAAGATTTTCAG--CGAGTTGCGCAGAGGGAGCTGTCTT 1347
QY 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
DB 1348 TCAGCCACCATGCTCTGAGGTGAGCATGAGATTTCTGAATAATTATGAGGCTCTCTGAG 1407
QY 382 GlnGluAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
DB 1408 CAGGAGAAATTAAGAAACAAATGCGGAGCTCTCTGTGATTCACCTATGATGTTGAT 1467
QY 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
DB 1468 GCGAAGCAAAAGACGATCAAGTTCATTAACATGAATGGGCTTAAGGAGCCCTATGAAA 1527
QY 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluIleThrPheArg 441
DB 1528 GTGTATAAAGATATGCGAGTTTATGAATGTTTGACATGCAACCATGAAAAGGAGATCTTAAAG 1587
QY 442 GlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnLysGln 461
DB 1588 GACAACTTTATTCAGATCCAAAACCTTGACATACTGATCACTTGGAGAGAGAG 1647
QY 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
DB 1648 AGGTGCTTCGATTTGTGTTTATTAATTAACCAAGAAAATGAGATTTAAAGCC 1707
QY 482 LeuValArgArgSerTyr--ArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1708 CTCGTCAAGAGGATTAATGAGAAACGACAGAGCAAAACCAAAATGCTTCACCTCG 1767
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1768 CAAAGAAAGAAAGTAAAGAAAGAAAGAG--GATTAAGACGAAAGAAACAGAAA 1821

```

```

QY 521 AspGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
DB 1822 AAAGAAAGAAAGAAAGATGAAGAGAGAAAGATGAAGAAAGAAAGCTCAAAAGAAAT 1881
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAsnAspGlu 560
DB 1882 ACCAAGAA-----AAGGACAGATAGATGGTACAGCA--GAGAAACTGAGGAA 1929
QY 561 LysGluAlaValAlaSerLysArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1930 AAGAGCAAGACCAACCCCGGGCGGAAAGACTGCGCAACAGTCAAGGCGCGGTAGAGGC 1989
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1990 CCGATCACCGAGTCCATGACAAAGCTGACCGAGCTGACCGACCGCAGCGCT 2049
QY 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
DB 2050 ACTGAAGACCCCAACCTCTGCCACCGCAACCAAGAACTTCTACAGAGCTGTG 2109
QY 611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
DB 2110 GAGACTCTCGATGAGCAAGAAAGAAATGGAAGTTGCTAAAAAGCTCATGAGAAAT 2169
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
DB 2170 GGTCTGTAATCGGACCAATTCCTAAATGATGGGAAACAAAGTCAACTCAATGTAA 2229
QY 651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
DB 2230 AACTCTATTTTAATCTATTAAGGCAACAAATCTTGACAACTCTTAACGACGCTTAA 2289
QY 671 LeuLysMetGluLysGluLysArgAsnAlaArgArgLysLysLysAlaProAlaAla 690
DB 2290 CAGAAAACTTCAAGAAACCTGCTGAAGAGCGAGATGTCTCAATGTAAAGTGCCT 2349
QY 691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluLysSerGlyVal 710
DB 2350 TCCACTGTCTCTGCT-----CAGAGAGATGAAGATTTGAAACCTCTC----- 2391
QY 711 SerGlyAsnGlnGlnGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
DB 2392 -----AATGAAGAAATAATCCAGAAAGCAGGAA----- 2421
QY 731 ValProArgGlyLysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
DB 2421 ----- 2421
QY 751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
DB 2422 -----GTTGAAGCTGTCAAG----- 2436
QY 771 AlaThrLeuGlyAlaAspGlyProProGlyProProThrProProArgTrpSer 790
DB 2437 -----CCACGAGAGACAGTCTGAAAAATGCTACTTCT 2469
QY 791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
DB 2470 CCGAGAAACACAGAACTCGGTGAGCTTGAG-----CCACACACGAGAACT 2517
QY 811 AlaPro--ProSerProSer--AlaProProProValValProLysGluGluLysGlu 828
DB 2518 GCAACCGATACATCTCTCTCTTAAGAGTTCCAAAGTACAAACCAAGCTGAAGATGAAGT 2577
QY 829 GluGluThrAlaAlaAlaProProVal----- 837
DB 2578 GTGGAGACCCAGGTGAATGACAGCATCACTGCTGAGACACGACAGCATGATGATGAT 2637
QY 838 -----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
DB 2638 CAGCAGAGACACAGTGTCAAGAGGTTCTGTGTGATGCCCAACCGCTACCAAA--- 2694
QY 852 AlaValAspThrGlyLysAlaGluGluProValLysSerGlyCysThrGluGluAlaGlu 871

```

Db	2635	---GCTGACTCTGTGAGCGTTGAAGTAGGGGTGCCAGAAACCATGACTTAAAGTTGAA	2751
Qy	872	GIUGIYProAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyLysLeu	891
Db	2752	GGTGATTAATACCAAGAAAGAGACTCTTGATAGGC-----	2787
Qy	882	LYbAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla	911
Db	2788	---ACTGAGAAAGGTGGAAACTTAGAATGAAAGATTGGTGATGCTCACCAGCAATAAATAGCC	2844
Qy	912	-----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal	926
Db	2845	CAGAGCCCGAGCCCCAGTCAGACACATAGATTCTCAGTGCAGCTGCAGCGCTGATGAG---	2901
Qy	927	ASpGluAlaGluGlyLysAspLysAsnArgLeuLeuSerPro-----ArgProSer	943
Db	2992	---GATGTGAGATGGAGACCAGACAGAGCGAGAGAAAGTTTCTCATGAGACTCAAGGCTTCA	2958
Qy	944	LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu	961
Db	2959	CTGTAAACCCCGACTGAGATCTATACTCGTC---TCATCTCGGTAAACCAATACCATG	3015
Qy	962	AspLeuLysGlnLeuLysGlnArgAlaAlaAlaAlaLeuProPheLe-	976
Db	3016	GATGTGCCACAGCTTCACAGCATGAGCTGTTATCCACCAATGATCTCTGCACCCCA	3079
Qy	977	-----GlnValThr	979
Db	3076	TGTAAACATACCAATTGGAACCCAGTAGCGGCTATGCTCTTACAGCGACACATTAA	3135
Qy	980	LYbValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro	999
Db	3136	GCAATGCATGAGTCAGCACTCTTGAG-----	3162
Qy	1000	ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro	1019
Db	3163	--GAGCAGCGCGCAGAGACAAGAACAGATTAATTGGATGATGAGAGTTCTTACAAAGTCA	3219
Qy	1020	ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAla	1039
Db	3220	TGTGGCACAATCCAAAGATGCA-----AACCAAGAG-----	3249
Qy	1040	GlnLysLeuProGlyAspProProCysThrPheSerGlyLeuProPheProValProPro	1059
Db	3250	-----TGC-----	3252
Qy	1060	ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro	1079
Db	3253	---GAAAGCTCTTACAGCTGCTCCACAT-----CAATTGATTAACATATCCCT	3297
Qy	1080	ProGlyHisProLeuLeuProLeuGlyLeuHisAspThrAlaAspProValLeuProArgPro	1099
Db	3298	GAAAGCGCTTGGGCTTCCG-----ACAACTCAGCAACACAGGCCACCGCC	3342
Qy	1100	ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu	1119
Db	3343	CCTCTCAATCCCGCTCATCCAAACACACAGCTGCTTCAGAAAA---CAATCTTTATA---	3396
Qy	1120	ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHis	1134
Db	3397	---ATGGAGAGCTCATCTCACAGGAAACACAGGCACTTATTGATCTTCATATACAG	3453
Qy	1135	ValProTyrSerGlnHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro	1153
Db	3454	GCTTCTTAATCTCAAGAAACACCCAGCGCTGATGAGATATCTCTTGGACCTGCA	3513
Qy	1154	LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSer	1173
Db	3514	CGGCAACAGGAATCTGCCAAATACGACTTTGCCCTTACATCAAGACAGAAAGATTTTCT	3573
Qy	1174	ProArgGlyGlnAlaGlyProProGlnSerLeuGlyValProThrAlaGlnGluAlaSer	1193

Db	3574	CCCCGAGCCAAACTCAAACTCGAGGGCTGTGGTGC	---AGGGCCCAACATGAAAGCT	3630
Qy	1194	ValleuATrgLYThralaleuGlySerValProGlyGlySerIleThrLYleGlyIlePro	1213	
Db	3631	GTAGTCAGAGGTAACCGCA	---GAGGCATACAAAGAGATATACCTCGGGAACTCCA	3687
Qy	1214	SerThralarValProSerApsSerAlaIleThrTYrarGlyGlySerIleThrHisGlyThr	1233	
Db	3688	ACCAACCAAAATTTTCAGTGGAGAGCATTCATCCCTCACGGGGCTCTATCATCTCAGGGCAC	3747	
Qy	1234	ProAla	-----ApyValleuTYrlvGlyIleThrIleThral	1245
Db	3748	CCGGCTCTCCGCCACATCGCATACCAACAGAGCTTTGGTAAGGGGTTCATTTCCGAA	3807	
Qy	1246	IleIleGlyIuApsSerProSerArGleuApsArGlyArGlyIuApsSerIeuProlys	1265	
Db	3808	ATGCCCATTTGAAGACAGCATGCT	-----GAGAAAGCCAGAGAGAACTGCATCCAA	3863
Qy	1266	GlyHisValIleTYrGlyIuLYleGlyHisValleuSerTYrGlyIuGlyIuMetSer	1285	
Db	3862	GGCCATGTTATTTATGAAGGCAGAAAGTGACATATCTTGATATGATATATTAAGAT	3921	
Qy	1286	ValThrGlyncysSerLYleuApsGlyArGlySerSerGlyProProHisGlyThrala	1305	
Db	3922	GCC	-----CGAGAAAGGACTAGAGTCCAGACACTGAAATACGT	3966
Qy	1306	AlaProLYsaTrgThrTYrApsMetMetGlyIuArGlyValGlyArGlyAlaIleSer	1323	
Db	3967	TTA	---AAGAAAGCTATGATCAGTGGAGAAATATTAAGCAAGGATGTCATGAG	4023
Qy	1324	-----SerAlaSerIleGlyIuLYleuMetGlyArGlyAlaIleProProGlyuArg	1339	
Db	4024	GAGTCTCCTGTATCAGACCGTTAGAGGGGCTGATATGCGAGCATTAACC	-----AGG	4077
Qy	1340	HisSerProHis	---HisLYleuGlyIuGlyHisIleValleuArgGlySerIleThrGlyIu	1358
Db	4078	GGGAGTCTCTATTTGACCTCAAGAAAGACTGATATGCTGGCTCATATACAGGG	4137	
Qy	1359	IleProArGlySerTYrValGlyAlaGlyIuApsTYrLeuArGlyAlaLYleu	1378	
Db	4138	ACACCAAGACACAACTGAAAGCTTTGAAAGATGGCTT	---AAATATCCAAACAAAT	4194
Qy	1379	LYsaTrgIuGlyThrProProProProProSerArGlyApsIleThrGlyIuLYleu	1398	
Db	4195	AAAAGGGAAAT	-----CTCTCCATACGAGCATTTGAAGTGGCAAT	4236
Qy	1399	ThrGlnAlaLeuGlyProLeuLYleuLYleuProAlaHisGlyIuLYleuValAlaThrVal	1418	
Db	4237	-----ACCAAAAGGAAACCA	---TATGATGGC	4272
Qy	1419	LYleuGlyAlaGlyArGlySerIleHisGlyIleProArGlyIuGlyIu	1433	
Db	4273	AAAAGAAATGGGGCTTCATTCATGAAATTCCAAGGCAAGATATTTAACTCAGAAAGT	4332	
Qy	1434	ArGlyHisThrProGlyIuLeuProLeuAlaProArGlyProLeuLYleuGlyIuSerIleThrGln	1453	
Db	4333	CGGAAATCTCCAGAAAGTGTCCAGAGCAACGGCCGCAATATTAAGGGTTCATTTCCAG	4392	
Qy	1454	GlyThrProLeuLYleuTYrApsThrGlyAlaIleSerThrThrGlySerLYleuHisApsVal	1473	
Db	4393	GGCACACCAATTAAGTTTGAACAACAC	---TCAGGTCAATCTGCATCAACAAACAATGTCT	4449
Qy	1474	ArgSerIleuIleGlySerProGlyArGlyThrPheProValHisProLeuApsValMet	1493	
Db	4450	AAATCTTAACTCAACGGGACCTTAGCAAACTATCCCGTGAATATGCTCGCTGAAATTTGTG	4509	
Qy	1494	AlaAps	---AlaArGlyAlaLeuGlyArGlyAlaCysTYrGlyIu	1507
Db	4510	CCAGAGAAACATTAAGTGTATGAAACGGGAAATATATGAGATGTGAAGACAGCGGAAAC	4569	
Qy	1508	LeuLYsaSerArGlyProGlyIuThralApsSerSerGlyLYleuSerIleAlaArgGlyAlaPro	1527	
Db	4570	GTGGTTCCTCCGGCACACGTAGTGGTAAAGTGTGGCCCTCGTCTTATAGTGCACA	---4626	

RESULT_15

1F087856

LOCUS	8018 bp	mRNA	linear	PRI 23-JUN-2003
DEFINITION	Homo sapiens nuclear receptor co-repressor mRNA, complete cds.			
ACCESSION	AF087856			
VERSION	AF087856.1			
KEYWORDS	GI:33150575			
SOURCE	Homo sapiens (human)			
ORANISM	Homo sapiens (human)			
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 8018)			
TITLE	Tu, Q., Yu, L., Dai, F., Yue, P., Huang, H. and Zhao, S. Y.			
JOURNAL	Cloning and characterization of a new human cDNA homologous to murine nuclear receptor co-repressor mRNA			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 8018)			
TITLE	Zhang, M., Yu, L., Zhou, Y., Hu, P. R., Xin, Y. R. and Zhao, S. Y.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (27-AUG-1998) Lab of Human Gene Research, Institute of Genetics, Pudan University, No. 220 Handan Rd., Shanghai 200433, P. R. China			
FEATURES	Location/Qualifiers			
source	1..8018			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	85..7116			
	/note="inchr"			
	/codon_start=1			
	/product="nuclear receptor co-repressor"			
	/protein_id="AAP97166.1"			
	/db_xref="GI:33150576"			
	/translation="MSSSVYPNQGAFSTEGQSNPSPHSVKTYTPSTTHQDDPAFGGNH EAPSSQPCGDDQDQNASPSKLSKEBLIOGMRVDREIAKEBQILKIKKQOOLEE EAKPPEPEKPVPEVPGHRSIVQIYIENKKEEAKIKFEGLGPYKELPLNPP NPPKXHEHNTKTNQWKRKLIFPKRNHAKROEKCORYOOLAMWSEKADYRIENR NPPKAKESKTRREYTERKQPEIKQREOQRFORQCRGLSGSLTARSHSEITD GISEQSNKQKQSLVTPPMFPAEORVYKFLNMGDLMEDPMKVTKDQRFMWVTDH EKKIFDKQFIQHPNFGFLASYLEKESVPCVLYITKYNENKALVRNYSGRGR NCOIAPPSQBEKYEKEEDKAEKTEKEKEKEDEKEDENKENTKDKIDGLEAE ETREBOATPRGRKTNASQGRKRGRTIRSTNTNAAAASAAATATEPPPLPPEP IETPEBTSWTEEMEWAKKGLVEHGRNOLAIAKMGTSNAQCKFYNYRNHL DNILQOKQKTSKRPRERDVSCESYASVYSAQDEDIANSBEENPESBGAENS DTSABSPVPEAVKPSDEPNATSRGNTEVLELPATYASTPSVLAIVSTKA EDSVETQVNDISASTAEQMDVDOOHESEBSGVLEPPATYASDVVRAVENNA SKYENGTKEERDLDRASEKYEPRDEDLVVAQINAPRQPSNDSEATCADYDVG EPRQNFPMDSKSLNPTGSLVSPKPNLDLPOLGRAVILPVMVSCPCNP ICRPVGYALYORHITKAMHESALREORORODIDLEBSTSPCGSKPMNEVYL QAPHOVITNLPFGVRLPTTRPTTRPPRLPLSSKTYVASKEPFIIMGGSIOGTEGY LTPNAAASYQETPKPSVGSISLGLPQOESASATLPLYKOEFSRSONSQPGLL VRAQHSVGRTGALIOEGSITRGTPSKISVSIISLRGSIIOGTPALQPTGIPTEA LVGSGISRMPIEDSSPEKGEBAASKGHVIYEGSGHILSYDNIKNAREGTRPRTAH EILKKSYSVEGNIKQGMRSRSPVASPLEGICALPGSPSHSLDKERTVLSGIM OCPRATTESFEDGLKTPKOIKRESPIPAFECALIKGKYDGIITIKMGRSHEIP KOILITQESKTPREVOSTRPIILEGSIQCTPIKFDNNSQSAIKHNVSLITPSKL SRMPLEIPIVENIKYVERKGYEDVAGAEVBRHRTSVSSGSAVSLTILAEAKQAL SPQIYDTSARTRPVSYONTMGRSGPMARISOVSSNKSNNHRSKSLTPTQESLPA KSPVGGDVVSHSPDPHHRGTAGVVRSHLPTHLDPMPHRAIDPAALVLPORA LSTPGVPSOYOLYAMENRQHTLINDIYISQOQVNLDPVAGLSAPRODLGAPYA TRQIDILTNMPTIILVPHRGISTPMDRITTYITGQITTPRPYPNYSANSGHPHTL AAASAREKER ETMLQORPSVFOGTNGTSVTLIPDPTAOLIMLPAPGSGISGLPASRNTADALAA ALVDAASAPQMDVSTKESKHAARLEENLRSAVASQOOLEKTEITVENRYSVOC LUNDAASAPSGKPPQHSVYSKAGKQKGPPEKRYEELERETKRTITANFIDVIT ROIASNDKAREGSSDSSSLSHRYEENPALTIVSASSAPROKLOITYOPEV VRANQANDPRTYRQEGPLHPRQESSPFOQLPPSSQAGMGQVPRTRILITLADH ICQITIDPRANQVSSQTPQOQPTSTFQNSPSALVSTPARKTSNYSPEQAQSVYH QRGSVSVSENLVDKSRGSPKSPRSHSVSSVPIISPPQVVAHEKDSILLLSQ RGAPLEQORNDAPSGSISYLPSPKFLNTSIPMYSKQKLEIRKUNSGGSGDMA ACGCTGIFNLPAVTSIGSVSRGSHPADPASNLGLDELIRKALMGSFDDKIVGVM SOMGVPGTANTSVVTSGETREEDDPHSGVCKDLIKSNRSKRSKPIIDGQGL GTRPSBSVSSYHSEGYHQTQPEMAEDNRPSSTGSGTQFPNPLTMMLSSTPTPLAC ABSAVQAQAHQOQNRIMWBRPAVLASQIETLSDSD"			
polya_signal	7984..7989			

polyA_signal

7984. .7985

ORIGIN

Alignment Scores:

Pred. No.:	1,396-73	Length:	8018
Score:	4043.00	Matches:	1061
Percent Similarity:	54.21%	Conservative:	330
Best Local Similarity:	41.35%	Mismatches:	731
Query Match:	30.59%	Indels:	444
DB:	9	Gaps:	104

US-09-522-753-5 (1-2517) x AF087856 (1-8018)

```

QY 137 LysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProProHis 155
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 190 CAGGATCCAGCATTCGAGGCAAAACATGAAGCTCCATCTCTCAATTCGCGGCAACCA 249
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 156 ThrAspProGluLeuGluLeuValProProArgLeuSerLysGluLeuLeuLeuLeu 175
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 250 TGTGAGATGATCAAAATGCTTCACCTTCACAAACCTCCAAAGGAGATTAAATACAGAGT 309
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 176 MetAspArgValAspArgGluLeuThrMetValGluGluGluLeuSerLysLeuLysLys 195
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 310 ATGATGCTGTGATGATGAGAAATTCGAAAAGTGAACAGCATCTTAAACTGAAAAAG 369
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 196 LysGluGluGluLeuGluGluGluValAlaLysProProGluProGluLysProValSer 215
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 370 AAACAACAACAGCTTGAAAGAGGAGCAGCTAAACCTCTGAGCCTGAGAGACCCGTGTC 429
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 216 ProProProLysLeuSerLysHisArgSerLeuValGluLeuLeuLeuLeuLeuLeu 235
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 430 CTTCTCTCTGTGAGAGAAACCCGAGATTGTCCAAATTATATGATGAGAAATCGG 489
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 236 LysLysValAGluAlaAlaHisArgLysLeuGluGluGluProGluProGluValGluLeuPro 255
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 490 AAAAAAGCAGAAAGACCTCTAAATTTTGAAGCTTGGCCCAAAAGTGAACCTGCCA 549
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 256 LeuTyraGluProSerAspThrArgGluTyraHisGluAsnLysLysLeuGluAla 275
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 550 CTGTATTAACAGCATCAGATACCAAGGTGTACATGAGAACATCAAGCAACCAAGGTG 609
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 276 MetArgLysLysLeuLeuLeuLeuTyraPheLysArgArgAsnHisAlaLysGlyGluTyrPhe 295
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 610 ATGAGGAAAAAACTCTTTTATTTTAAAAAGAAATCATCATCAAGAAAAAACAAGGAA 669
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 296 GlnLysPheCysGluGluGluTyraAspGluLeuMetGluAlaLeuGluLysLysValGluArg 315
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 670 CAATAAATCTGCCAGCGTTATGATCAGCTCATGAGAGCATGGAGAAAAAGTGGACAGA 729
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 316 IleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrgluLys 335
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 730 ATGAGAAATATATCTCTCGAGAGAAAGCTTAAAGAAAGCAAAACAAGGAAATCTATGAAAAG 789
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 336 GlnPheProGluLeuArgLysGluArgGluLeuGluGluGluGluMetGlnSerArgValGly 355
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 790 CAGTTTCCAGAAATTCGAAAAACAAGAACACCAAGAAAGATTTGAG---CAGTTGGG 846
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 356 GlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluLeu 375
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 847 CAGAGGGAGACTGCTTTTCAAGCCACCATTTGCTAGAGTGAAGCATGAGATTTCTGAAATT 906
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 376 IleAspGlyLeuSerGluGluGluGluGluLeuGluGluGluMetArgGluLeuAlaValIle 395
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 907 ATTGATGGGCTCTCTGAGCAGGAAGTAATGAGAAACAATTCGGCGAGCTCTCTGTGATT 966
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 396 ProProMetLeuTyraAspAlaAspGluGluArgLysLysPheIleAsnMetAsnGlyLeu 415
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 967 CCACTTATGATGTTTATGACGAACAAGCAGATCAAGTTCAATTAAACATGAATGGGCTT 1026
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 416 MetAlaAspProMetLysValTyrlsAspArgGluValMetAsnMetTyrSerGluGln 435
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1027 ATGAGAGACCTTATGAAAGTGTATTAAGATAGGCACTTTATGATGATTTTGACATGACAT 1086
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 436 GlnLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAla 455

```

```

DB 1087 GAAAAAGAGATCTTTAAGACAAAGTTTATCCAGCATCCAAAAAACTTGGACTAATTGCA 1146
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 456 SerPheLeuGluArgLysThrValAlaGluCysValLysTyraTyraLeuThrLysLys 475
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1147 TCATCTTGAGAGAGAGAGAGTCTCTGATGTGTGTTTGTATTCATATTAAACAAAGAA 1206
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 476 AsnGluAsnTyrlsSerSerLeuValArgArgSerTyraArgArgGlyLysSerGln 494
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1207 AATGAGAAATTATTAAGCCCTGCTCAGAAAGAAATTAAGGAAAGCAGAGAGCAAGAACAG 1266
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 495 GlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1267 CAATATTGCTCGACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 515 SerSerGlnGluGluLysAspGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGly 534
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1321 GCAGAAAAAACAAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 535 LysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSer 554
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1381 GAAGACTCCAAAGAAAAATACCAAGGAA-----AAGAGACAAGATAGATGATCAGCA 1431
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 555 GlyLysAspAsnAspGlyLysGluValAlaValAserLysGlyArgGlyThrAlaAsnSer 574
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1432 ---GAAGAAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 575 GlnGlyArgArgLysGlyArgLysThrArgSerMetAlaAsnGlyLysAsnSerGluGln 594
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1489 CAGGCGCCGCTGAAGGCGGAGTCAACAGTCCATGACAAAGAGAGCTCAGCTGCAGCT 1548
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 595 AlaIleThrProGlnGlnSer-----AlaGluLeu 604
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1549 GCTGCGAGCCGACGCGCTGTGAAGAGCCCAACCACTCTGCCACCGCACAGAGAGAGCC 1608
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 605 AlaSerMetGluLeuAsnGlnSerAspArgTyrThrGluGluGluMetGluThrAlaLys 624
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1609 ATTTCTACAGACCTGTGAGACCTTCGATGACAGAAAGAAATGGAAGTGTCTAATA 1668
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 625 LysGlyLeuLeuGluHisGlyArgAsnTyrPheSerAlaAlaArgMetValGlySerLys 644
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1669 AAGAGCTTGAAGACATGCTGTAACTGTAACCTGGCAGCAATTCCTAAATGCTGGAGAGAA 1728
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 645 ThrValSerGlnCysLysAsnPheTyraPheAsnTyrlsLysArgGluAsnLeuAspGlu 664
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1729 AGTGAAGCTCAATGTAAATCTTCTATTTAATTAAGAGAGAGACAAATCTTGAACAAC 1788
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 665 IleLeuGlnGlnHisLysLeuLysMetGlyLysGluArgAsnAlaArgArgLysLys 684
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1789 CTCTTACACACACATTAACAGAAACCTTCACAGAAACCTCGTGAAGAGAGAGATGTCTCT 1848
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 685 LysAlaProAlaAlaAserGluGluAlaAlaPheProProValAlaGluAspGluGln 704
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1849 CAATCGAAAGTGTCTCTCCACTGTTCTGCT-----CAGAGAGATGAAGAT 1896
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 705 MetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeu 724
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1897 ATTGAAGCTTCC-----AATGAAGAAAGAAATCCAGAGAGACGAGAGAGCT--- 1941
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 725 HisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsn 744
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1942 -----GCAAGAAAT 1950
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 745 SerSerAspThrGluSerLysProSerProHis---ThrGluAlaAlaLysAspThrGly 763
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1951 AGTTCTGATACAGAAAGTGTCTCTTCTCTTCAACAGTTGAAGCTTCAAG----- 2001
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 764 GlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProPro 783
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2002 -----CCACAGCAGAGAC 2013
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 784 ThrProProArgArgThrSerArgAlaProLysGluProThrProAlaSerGluAlaThr 803
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db	2014	AGTCTGAAAATGCTACTTCTCGAGAAAACACAGAACTCGCGGTTGAGCTTGAG-----	206
Qy	804	GLyAlaProThrProProProAlaPro---ProSerProSer---AlaProProProVal	821
Db	2068	-----CCGACCGAGGAAACCTGCAACCCAGTACATCTCCCTCTTACGAGTTCCAAAGTCA	2121
Qy	822	ValProLysGluGluLysGluGluGluThrAlaAlaAlaProProVal-----	837
Db	2122	AAACCAAGCTGAAGTGAAGAAAGTGGAGACCCGAGTAATGACAGCATGCTGTGAGACA	2181
Qy	838	-----GluGluGluGluGluGlnLys	844
Db	2182	GCAGAGCCAGATGATGTGATGCAGACGACACTGCTGAAGAGGGTTCTGTTTGAT	2241
Qy	845	ProProAlaAlaGluGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSer	864
Db	2242	CCCCACCCCGGCTACCAA-----GCTGACTCTGACAGCTTGAAGTAGAGGTGGCACA	2295
Qy	865	GluCyThrTrgGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGlu	884
Db	2296	AACCATGCATCTTAAAGTTGAAGGTGATATATCAAAAGAAAGAGACTTGATGATAGACC---	2352
Qy	885	AlaThrAlaGluGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr	904
Db	2353	-----AGTGAAGAGGTGAACCTAGAGATGAAGATTGGTG	2388
Qy	905	ThrAlaLysSerSerGlyAla-----ProGln---AspSerAspSerSerAla	919
Db	2389	GTAGCTCAGCAATTAATATGCCAAAGGCCGAGCCCGACCTGTCAGACATGATTTCCAGTCC	2448
Qy	920	ThrCySerAlaAspGluValAspGluAlaGluGlyGlyAspLysAspAspLeuLeuSer	939
Db	2449	ACGTGCACGCGTGAATGAG-----GATGTGATGGAGAGCCAGAGAGCGAGAGATGTTT	2502
Qy	940	Pro-----ArgProSerLeuLeuThrProThrGlyAspProAlaAlaAspLase	956
Db	2503	CCTATGCACTCAAAAGCTTCACTGATTAAACCCCACTGGATCTATATCTGTC---TCACCT	2559
Qy	957	ProGlnLys-----ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaPro	974
Db	2560	CCGTTTAAACCAATTCACGTGATCTGCCACAGCTTCAGCATCGAGCTGCTGTATCCCA	2619
Qy	975	ProIle-----	976
Db	2620	CCAAATGGATCTGCAACCCCATGTAACTACCAATTGGAACCCAGTAGGGCGTATGCT	2679
Qy	977	-----GlnValThrLysValHisGluProProArgGluAspAlaAlaProThr	992
Db	2680	CTTCAACAGCAGACATTAAGCAATGATGATGACAGCATCTCTGGAG-----	2727
Qy	993	LysProAlaProProAlaProProProGlnAsnLeuGlnProGlnSerAspAlaPro	1012
Db	2728	-----GAGCAGCGGCGAGACAGACAAGAACATAGATTGGAA	2763
Qy	1013	GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLys	1032
Db	2764	TGTGAAGATTACAAAGTCCATGTGGCACAATCCAAAGTCCA-----AACAGA	2811
Qy	1033	GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCyThrThrSerGly	1052
Db	2812	GAG-----TGG-----	2817
Qy	1053	LeuProPheProValProProArgGluValLleLysAlaSerProHisAlaProAspPro	1072
Db	2818	-----GAACTCTTACAGCTGCTCCACAT-----	2841
Qy	1073	SerAlaPheSerGlyAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla	1092
Db	2842	CAAGTGAATACTAATCTCCCTGAAGGCGTTGGCTCG-----ACAAC	2886
Qy	1093	ArgProValLeuProArgProProThrLleSerAsnProProLeuLleSerSerAla	1112
Db	2887	CGACCAACCAAGGCCACCGCCCTCTCTATCCCGCATCAAAACCAACGAGCTTGACGAA	2946

QY	1113	LYVHISPR	SErValLeuGluArgGlnIleGlyAlaIleSerGlnGly	-----	1128
Db	2947	AAA----	CCATCTTTTATA-----ATGGAGAGCTCCATCTTCACAGGGGAACACGAGCACT	2997	
QY	1129	---MetSerVal	GlnLeuHleValProTyrSerGlnHisAlaIleValAlaPro---ValGlyY	1146	
Db	2998	TATTTGACTTCTCATTAATCAGAGCTTCTTCACTCAAGAAACACCCAAAGCCCTCACTGGGA	3057		
QY	1147	ProValThm	MetGlyLeuProLeuProMetAspProLysIleValAlaProPheSerGly	1166	
Db	3058	TCATCTCTCTTGGAGCTGCCACGGGCAACAGAAATCTGCCAAATACGACTACTTCCCTTAC	3117		
QY	1167	ValIysGlnGlnGlnLeuSerProArgGlyGlnAlaGlyProProGlnSerLeuGlyVal	1186		
Db	3118	ATCAAGCAGGAAGAAATTTTCTCTCCCGAAGCCAAACTCACACCTGAGGGCTCTTGGCTC	3177		
QY	1187	ProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGly	1206		
Db	3178	---AGGGCCCAACATGAAGGTGTAGTCAGAGGTACCGCA---GGAGCATATCAAGAAAGA	3231		
QY	1207	SerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArg	1226		
Db	3232	AGTAACTCTCGGGAACTCCAAACAGCAAAATTTCACTGGAGACATTCATCCCTCCG	3291		
QY	1227	GlySerIleThrHisGlyIleThrProAla-----AspValLeu	1238		
Db	3292	GAGCTTATCACTACAGGGACCCCGGCTCTGCGCCAGACTGGCATACCAAGAGCTTTG	3351		
QY	1239	TyrIysGlyThrIleThrArgIleIleGlyGlnAspSerProSerArgLeuAspArgGly	1258		
Db	3352	GTGAAGGGGTCATTTGAGAAATCCCATTTGAAGACAGCACTCT-----GAGAAAGGC	3405		
QY	1259	ArgGlnAspSerLeuProLysGlyHisValIleTyrGlnGlyLysGlyHisValLeu	1278		
Db	3406	AGAGAGGAAGTCGATCCAAAGSCCATGTTATTTATAGGCAAAAGTGCACATATCTTG	3465		
QY	1279	SerTyrGlnGlyGlyMetSerValThrGlnCysSerLysGlnAspGlyArgSerSerSer	1298		
Db	3466	TCATATGATATATTTAAGAATGCC-----CGAAGAAAGGACTAAGAGTCCA	3510		
QY	1299	GlyProProHisGlyThrAlaAlaIleProLysArgThrTyrAspMetMetGlnGlyArgVal	1318		
Db	3511	AGAACAGCTCTTGAAATCAGTTTA---AAGAAAGCATATGATACGTGAAGGAATATA	3567		
QY	1319	GlyArgAlaIleSer-----SerAlaSerIleGlnGlyLeuMetGly	1332		
Db	3568	AAGCAAGGAGATGTCAATAGAGGAATCTCTGTATCAGACCCGTTAAGAGGGCTATATGC	3627		
QY	1333	ArgAlaIleProProGlnArgHisSerProHis---HisLeuLysGlnGlnHisIle	1351		
Db	3628	CGAGCATTAACCC-----AGGGGAGTCTCTATCTGACCTCAAAAGAAAGACTGTATTG	3681		
QY	1352	ArgGlySerIleThrGlnGlyIleProArgSerTyrValGlnAlaGlnGlnArgTyrLeu	1371		
Db	3682	TCCTGCTCCATATAGCAGGGGACACCAAGAGCAACATGAAGCTTTGAAGAGGCTT	3741		
QY	1372	ArgArgGlnAlaLysLeuLeuLysValArgGlyGlyThrProProProProProSerArg	1391		
Db	3742	---AAATATCCCAACAAATTTAAAGGAAAGT-----CCTCCCATACGA	3783		
QY	1392	AspLeuThrGlnAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis	1411		
Db	3784	GCATTTGAGGTGCATT-----ACCAAAAGAAACCA---TAT	3819		
QY	1412	GlnGlyLeuValAlaThrValLysGlnAlaGlyArgSerIleHisGlnIleProArgGly	1431		
Db	3820	GATGGC---ATCACCAACATCAAGAAATGGGGGCTTCCATTCATGATTTCCAAAGGCA	3876		
QY	1432	GlnLeu-----ArgHisThrProGlnLeuProLeuAlaProArgProLeu	1446		
Db	3877	GATATTTTAACTACGAAGAGTCGGAATAATCTCCAGAGTGTCTCCAGAGCAACACGGCCGAT	3936		

QY 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
 DB 3937 ATTTGAGGCTTCCATTTCCACAGGGCACCAATATAAGTTTGACAACAC---TCAGTCA 3993
 QY 1467 GlySerLysLysHisAspValArgSerLeuIleGlySerProGlyLysThrPheProPro 1486
 DB 3994 TCTGCATCAAAACACATGTCAATCTTAATCACGGGGCTTAGCAAACTATCCCTGGA 4053
 QY 1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGlnLysArgLysCysTyrGlu 1505
 DB 4054 ATGGCTCCCTGGAAATTTGTGCCAGAAACATTAAGTGTAGAACGGGGAATAATATGAG 4113
 QY 1506 -----GluSerLeuLysSerArgProGlyThrAlaSerSerGlyGly 1520
 DB 4114 GATGTAAAGCAGGCGAGACCGTGCCTTCCGSCACACGTCAAGTGTATGCTCTGGCCCC 4173
 QY 1521 SerIleAlaArgGlyAlaProValIleValProGlnLeuGlyLysProArgLysProPro 1540
 DB 4174 TCCGTTCTTAGGTCACA-----CTGCATGAAGCTCCCAAAACACACTGAGCCCT 4224
 QY 1541 LeuThrTyrGluAspHisGlyAla-----ProPheAlaGlyHisLeuPro 1555
 DB 4225 GGGATTATGATGACCCAGTGCACGAGAGACCCCTGTGATGTTATAAAACACCATGTC 4284
 QY 1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSer 1575
 DB 4285 AAGAGCTCAACCATGATGAACAGA-----ACTTGATGATGTTCT 4323
 QY 1576 SerSerLysAla---SerGlnAspArgLysLeuThrSerThrProArgGlu----- 1591
 DB 4324 TCTAACAAAGTCTACCAATCATGAAAGAAATGCACACTGACCCCTCCAGAGGGAAAGT 4383
 QY 1592 ---IleAlaLysSerProHisSerThrValProGlnHisIleProHisProLieserPro 1610
 DB 4384 ATCCACGACGAGTCTCAGTGCCTGGGGTGGACCTGTGTGAGCCAC-----AGTCCG 4437
 QY 1611 TyrGlnHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeu 1630
 DB 4438 TTTGATCCCATCAGACGAGGACACTGCAGGAGAGTTTATGGAGCCACCTGCCACAG 4497
 QY 1631 AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleTyr 1650
 DB 4498 CACTTGATGCCA---GCCATGCTTTTTCACAGGGCTTTGATCCTCGGTGCTTACTCG 4554
 QY 1651 LeuProArgHisLeuAlaProAspProThrTyrProHisLeuTyrProProTyrLeuIle 1670
 DB 4555 TTTCAGAGACGCTTCCCACTCCAGGTTACCCAAAGTCAGTATCAGCTTAC----- 4608
 QY 1671 ArgGlyTyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyr 1689
 DB 4609 -----GCCATGGAGAACACAAAGACAGACATCTTAATATGATTAC 4647
 QY 1690 IleThrSerGlnGlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
 DB 4648 ATTACCTCCACACAGATGCAAGTGAACCTTG-----CCTCCAGATGTG 4689
 QY 1710 LeuArgGlyLeuSerProArgGlyLeuSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
 DB 4690 GCCAGAGGACTCTCCCAAGAGAGACCACTGGGCTCCCATACCCAGCA---ACGAGA 4746
 QY 1730 GlyIleIleAspLeuSerGlnValProHisLeuProValLeuValAProProThrProGly 1749
 DB 4747 GGAATATTGACTGACCAATATGCGCTCA---ACAATTTTACTGCTCATTCAGGGGGA 4803
 QY 1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
 DB 4804 ACAAGCACTCTCCCAATGACAGAAATCACTTATATCTCGTACACAGATTAATCTTCCCT 4863
 QY 1770 SerArg---HisSerSerProLeuSerProGlyLysProThrHisLeuThrLysPro 1788
 DB 4864 CCCAGGCGGTCAACTCTGCTTCATGTCCTCCAGAGACCCAAACACCTT----- 4914
 QY 1789 ThrThrThrSerSerSerGluArgLysArgAspArgAspArgGluArgAspArgAspArg 1808

DB 4915 GCAGCTGCTGCAAGTGTGAGAGGGAGACGGAAACCGGAGCGGAGAGAGCGG 4974
 QY 1809 GlnArgGluLysSerIleLeuThrSerThrThrValGlnHisAlaProIleTyrArg 1828
 DB 4975 GAACGG-----ATTGCTCAGCTTCTCCAGCTC-----TACCTGGG 5013
 QY 1829 ProGlyThrGlnGlnSerSerGlySerSerGlyGlyGlyGlyLysSer 1848
 DB 5014 CCAGGCTCAGAAACAG-----CCT 5031
 QY 1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProLieserProArgThrGln 1868
 DB 5032 GGCCGACCTGGCAGTCAATGATGTGCTCCCTCCCTCCCT---TCAGTGAAGACTGAG 5088
 QY 1869 AspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIle 1887
 DB 5089 GAGACCAATGTTGCAACAGAGACCCAGTGTTTCCAAAGAACCAATGAAACAGTATATC 5148
 QY 1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerProVal 1907
 DB 5149 ACACTTTGATCCAACTGCTCAGTCAAGATCATGCCACTGCTGGGGGCTTCA 5208
 QY 1908 -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
 DB 5209 ATAAAGCAAGGCTGCGACCTCCCGTATCAACACTGCTGCGAGATGC---CTGGCTGCT 5265
 QY 1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys----- 1940
 DB 5266 CTGTGGAGTGTGTCAGCTTCTGACCCCAAGATGATGTGTCCAAACAAAGAGAGTAAAG 5325
 QY 1941 ---GlnAlaProArgValAlaArgProGluArgProArgAlaAsp----- 1954
 DB 5326 CATGAAGCTGCAGGTTAGAAAGAAATTTGAGAGACAGGTGACAGACTAGTAGACAG 5385
 QY 1955 -----Thr 1955
 DB 5386 CAGCACTAGACAGAAACCTGAGTGGAGAAAGATCTGTCACTGTTATACACT 5445
 QY 1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGlnProAlaSerSer--- 1974
 DB 5446 TCTTCAAGCTTTCAGATGCGCAAGGCC-----CAGCTCATCTTTCAGTA 5490
 QY 1975 -----ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla 1992
 DB 5491 GTTATTTGAGAGTGGGAAAGATTAAGGCT-----CTCCAAATCCAGATATGAG 5544
 QY 1993 ThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisIleAlaSerProAsp 2010
 DB 5545 GAAGAGCTTAAGGACAGAGAGAACTACATTAAGTCAAGCTTAAGTCAAGAGCTGATC 5604
 QY 2011 ProProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysPro 2030
 DB 5605 ATCACCCTGGAATTTGCTCGAGCAAGATGCGAGAGAACTGGCTCTAAAGTTCAAGC 5664
 QY 2031 PheSerIleGlnGlnLeuGlnLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
 DB 5665 TCTTCT-----AGTACTATGCTTTCACAGATGATAACACTAGC 5706
 QY 2051 GlnGlyValGluProValSerProValSerSerProSerLeuThrHisAspArgGlyLeu 2070
 DB 5707 GATGCTATTGAGATGATTAAGTCTGTCAGACTACCTGCGSCCAACCCAGAGAAATGTCAG 5766
 QY 2071 ProLysHisLeuGlnGlnLeuAspLysSerHisLeuGlnGlyGlnLeuArgProLysGln 2090
 DB 5767 ACCTATCAGCCAGAGTGTGTTAAGGCMAATCAAGCGGAAAGATGCTTCAAGACATAT 5826
 QY 2091 ProGlyProValLysLeuGlyGlyGlnAlaAlaHisLeuProHisLeuArgProLeuPro 2110
 DB 5827 GAAGAGCA-----TTACATCACTATGAGCA----- 5853
 QY 2111 GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly----- 2125

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 11, 2005, 09:34:57 ; Search time 24239 Seconds
(without alignments)
3952.629 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPRKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.p2n.model -DEV=x1p
-Q=/cgn2.1/USPTO.spool_p/US09522753/runat_02112005_173640_20807/app.query.faeta-1.2695
-DB=EST -QFMT=faetap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753 @CGN 1.1.10059 @runat_02112005_173640_20807 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONJLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8226	62.2	7372	9	AY412686 Homo sapi
2	7109.5	53.8	7013	9	AY412688 Mus muscu
3	6533	49.4	6836	9	AY412687 Pan trogl
4	2566	19.4	2000	3	BC033087 Homo sapi
5	2264	17.1	1362	3	BC020427 Homo sapi
6	1643.5	12.4	1966	3	BC026028 Homo sapi
7	1634	12.4	1850	3	BC035748 Homo sapi
8	1626.5	12.3	1835	3	BC029627 Mus muscu
9	1624.5	12.3	1819	3	BC062162 Mus muscu

10	1624	12.3	1093	5	BM909096	BM909096 AGENCOURT
11	1603	12.1	1996	3	BC026623	BC026623 Mus muscu
12	1559	11.8	1965	4	BM555371	BM555371 AGENCOURT
13	1540	11.7	1126	4	BM553310	BM553310 AGENCOURT
14	1480.5	11.2	1057	4	BM477568	BM477568 AGENCOURT
15	1479.5	11.2	1051	4	BM423558	BM423558 AGENCOURT
16	1471.5	11.1	937	5	BQ949384	BQ949384 AGENCOURT
17	1448	11.0	846	5	BQ957144	BQ957144 AGENCOURT
18	1442.5	10.9	984	5	BQ690869	BQ690869 AGENCOURT
19	1440.5	10.9	1159	4	BM802749	BM802749 AGENCOURT
20	1420	10.7	1022	5	BM910785	BM910785 AGENCOURT
21	1382.5	10.5	1030	4	BM558844	BM558844 AGENCOURT
22	1378.5	10.4	1010	5	BM915686	BM915686 AGENCOURT
23	1374	10.4	887	5	BQ542258	BQ542258 AGENCOURT
24	1368	10.4	875	7	CK769773	CK769773 957873 MA
25	1366	10.4	1067	4	BM472005	BM472005 AGENCOURT
26	1366.5	10.3	868	5	BX368971	BX368971 AGENCOURT
27	1348.5	10.2	923	5	BU184403	BU184403 AGENCOURT
28	1346.5	10.2	1038	5	BM910704	BM910704 AGENCOURT
29	1338	10.1	887	6	CA979881	CA979881 AGENCOURT
30	1334	10.1	923	5	BX390462	BX390462 AGENCOURT
31	1332	10.1	875	5	BQ711119	BQ711119 AGENCOURT
32	1331.5	10.1	842	5	BU224569	BU224569 603400211
33	1326	10.0	1094	4	BM560912	BM560912 AGENCOURT
34	1325.5	10.0	905	5	BQ897825	BQ897825 AGENCOURT
35	1321	10.0	882	5	BU180236	BU180236 AGENCOURT
36	1305.5	9.9	898	5	BU172348	BU172348 AGENCOURT
37	1305	9.9	1045	4	BM560255	BM560255 AGENCOURT
38	1297.5	9.8	833	7	CK778015	CK778015 96516 MA
39	1274	9.6	875	5	BQ691710	BQ691710 AGENCOURT
40	1272	9.6	3575	3	AL137641	AL137641 Homo sapi
41	1267	9.6	769	5	BX368972	BX368972 AGENCOURT
42	1262.5	9.6	915	7	CF272442	CF272442 AGENCOURT
43	1260	9.5	1075	4	BM451469	BM451469 AGENCOURT
44	1259.5	9.5	785	7	CO896175	CO896175 Boven 24
45	1256.5	9.5	928	5	BU164114	BU164114 AGENCOURT

ALIGNMENTS

RESULT 1	AY412686	7372 bp	DNA	linear	GSS 16-DEC-2003
LOCUS	AY412686				
DEFINITION	Homo sapiens NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY412686				
VERSION	AY412686.1	GI:39768651			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M. (bases 1 to 7372)				
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Infering nonneutral evolution from human-chimp-mouse orthologous gene titios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
REFERENCE	2 (bases 1 to 7372)				
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
SOURCE	1..7372				

gene
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>7372
/gene="NCOR2"
/locus_tag="HCM4640"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	7372
Score:	8326.00	Matches:	1708
Percent Similarity:	69.56%	Conservative:	1
Best Local Similarity:	69.52%	Mismatches:	748
Query Match:	62.25%	Indels:	0
DB:	9	Gaps:	0

US-09-522-753-5 (1-2517) x AY412686 (1-7372)

```
QY      1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
        |||
DB      1 ATGTGGGATTCACACAGCCTGTGCACAGACGTGGAGGGCCACTGAGCCCGCTACCG 60
QY      21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrArgValGlyLeuLeu 40
        |||
DB      61 CCCCAAGCCCTTCTTACCCAGTGCAGATGCCGAGACGACACGACGCGTGGGCTCTG 120
QY      41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
        |||
DB      121 GAGTACCAACACACACTCCGCGACATATGCTCCCACTGTCCGCGGCTCCATCAACCG 180
QY      61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValangluArgSerGln 80
        |||
DB      181 CCCCAAGCCGCGAGGCGCTCCCTGCTGTGTGAGTTCCAGCCCGGAGATGAACGNNNNNN 240
QY      81 GuluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGluLysSerGluMet 100
        |||
DB      241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300
QY      101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProArgProLeuLeuArgPro 120
        |||
DB      301 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360
QY      121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
        |||
DB      361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
QY      141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrArgProGluLeu 160
        |||
DB      421 CTGACGGGCAAGCTGGACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCTTGAGCTG 480
QY      161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
        |||
DB      481 GAGCTGTGCTCCGCGACGCTGTCCAAAGAGAGACTATCCAGAACATTCGACCGCGTGCAG 540
QY      181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
        |||
DB      541 CCAAGAGATCAACCATGTAGAGCAGCAGATCTTAACCTGAAGAAAGACGACACAGCTG 600
QY      201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
        |||
DB      601 GAGAGAGAGAGCTGCCAAGCGCGCGAGCCTGAAGAACCCGCTGCACCGCGCCCATCGAG 660
QY      221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
        |||
DB      661 TCGAAGCACCGCAGCTGTGTGCAGATCATCTACGAGAGAACCGGNNNNNNNNNNNNNN 720
QY      241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGlnPro 260
        |||
DB      721 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 780
QY      261 SerArgSerThrArgGlnTyrHisGluAsnIleLysValLeuGlnAlaLeuLysLysLeu 280
        |||
DB      781 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACAGGCGATGCGGAAGAGCTA 840
```

```
QY      281 IleLeuTyrPheLysArgArgHisHisAlaArgLysGlnTrpLysGluLysPheCysGln 300
        |||
DB      841 ATCTTGTAATCTCAAGAGAGAAATCACGCTCGAAACAAATGGAGACGAAGTTCTGCCAG 900
QY      301 ArgTyrArgGlnLeuMetGluAlaLeuGluLysLysValGluAlaArgIleGluLysAsnPro 320
        |||
DB      901 CGCTATGACACACTCATGAGAGGCTGGGAGAAAGAGTGAGCGCATTCAGAAACAACCCC 960
QY      321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
        |||
DB      961 CGCGCGCGCGCCAAAGAGAGCAAGGTGCCGAGTACTAGAGAGAGATTCCTTGATATC 1020
QY      341 ArgLysGlnArgGluLeuGlnGluLysArgMetGlnSerArgValGlyGlnArgLysSerGly 360
        |||
DB      1021 CCAAGCAGCGCAGAGCTGCAGAGCGCATGACAGACGAGGTGGGCCACGCGGCGAGTGGG 1080
QY      361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
        |||
DB      1081 CTGTCCATGTCCGCGCGCCGCGAGCAGCAGAGGTGTCAAGATCATCGATGGCCTCTCA 1140
QY      381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
        |||
DB      1141 GAGCAGAGAACTTGGAGAGACAGATGCCCAAGCTGTGATCCCGCCCATGCTGTAC 1200
QY      401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
        |||
DB      1201 GACGCTGACCAAGCGCATCAAGTTCAATCAATGAACGAGCTTATGGCCGACCCCATG 1260
QY      421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
        |||
DB      1261 AAGGTGTACAAAGACCGCAGGTCAATGAACATGTGAGAGACACAGAGAAAGACCTTC 1320
QY      441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
        |||
DB      1321 CCGGAGAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1380
QY      461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
        |||
DB      1381 NNNACAGTGGCTGAGTGCCTCTTACTACTACGACTGAAGAGATGAGAACTATAG 1440
QY      481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
        |||
DB      1441 AGCTGTGTGAGACGAGAGCTATGCGCGCCGCGCAAGAGCACAAGNNNNNNNNNNNN 1500
QY      501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
        |||
DB      1501 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
QY      521 AspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsn 540
        |||
DB      1561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
QY      541 AspLysGluAspLeuLeuLysGluLysThrAspArgThrSerGlyGluAspAsnAspGlu 560
        |||
DB      1621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1680
QY      561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
        |||
DB      1681 AAGAGAGCTGTGCTCCAAAGCCCGCAAAATGCGCAACAGCAGGAGAGAGCGCAAAAGC 1740
QY      581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
        |||
DB      1741 CCGATCACCCGCTCATATGAGGCGCAACAGAGAGAGGCCATTCACCCCGCAGCAG 1800
QY      601 SerAlaGluLeuValAsnMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
        |||
DB      1801 AGCGCGAGCTGCTCCATGTAGCTGAATGAGAGTTCGCTGTGACAAAGAGAAAG 1860
QY      621 GluThrAlaLysLysGluLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
        |||
DB      1861 GAAACAGCCAAAGAAAGTCTCTTGAAACACGCGCCGCACTGTGCGCCATTCGCGCGGATG 1920
```

QY 641 ValGlySerIysThrValSerGlnCysIysAsnPheTyrPheAsnTyrLysValArgGln 660
Db 1921 GTGGGCTCCAAAGACTGTGTGCGACTGTAAAGAACTTCTACTTCACTCAAGAAAGAGCGAG 1980
QY 661 AsnLeuAspGluIleuGlnGlnHisIleuValMetGluLysGluValArgAsnAlaArg 680
Db 1981 AACTCCATGAGATCTTGACAGACACAAAGCTGAAGATGAGAAAGAGAGAGAGCGCGG 2040
QY 681 ArgIysIysLysValAspAlaProAlaAlaAspSerGluValAlaPheProValVal 700
Db 2041 AGGAAAGAAAGAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2100
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2101 GAGGATGAGAGAGATGAGAGCGCTCGGCGTGAGCGAAATGAGAGAGAGATGTGTGAGAG 2160
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2161 GCTGAAGNN 2220
QY 741 ThrValAsnAspSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2221 NNN 2280
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2281 NNN 2340
QY 781 GlyProProThrProProAlaArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2341 NNN 2400
QY 801 GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2401 NNN 2460
QY 821 ValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGly 840
Db 2461 NNN 2520
QY 841 GluGluGluLysProProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu 860
Db 2521 GAGAGAGAGAGAGCGG 2580
QY 861 ProValIysSerGluCysThrGluGluAlaGluGluGluProAlaLysGlyLysAspAla 880
Db 2581 CCGCTCAAGAGCGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
QY 881 GluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyGlySer 900
Db 2641 GAGGCGGCTGAGGCGAGCGCGCGAGGGGGCGCTCAAGCGCAAGAAAGAGGGGCGGAGC 2700
QY 901 GlyAlaArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Db 2701 GCGAGGCGCCACACAGCGCAAGAGCTCGGGCGCGCGCGCGCGCGCGCGCTCAGTGTACC 2760
QY 921 CysSerAlaAspGluValAspGluAlaGluGlyGlyValAspLysAsnAlaGluLeuSerPro 940
Db 2761 TGCAGTGCAGACCAAGTGTAGTGAAGCGCGAGCGCGCGCGCGCGCGCGCTGTCTCCCA 2820
QY 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro 960
Db 2821 AGGCCAGAGCTCTCAAGCGCGAGCTGGGCGAGCGCGGGCGCAATGCTGTCAAGCGAGGCA 2880
QY 961 LeuAspLeuLysGlnLeuLysGlnAlaAlaAlaIleProProIleGlnValThrLys 980
Db 2881 CTGGAGCTGAAGAGCTGAAGAGCGAGCGAGCGAGCTGCCATCCCGCGCAACGAGTCAACCAA 2940
QY 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Db 2941 GTCCATGAAGCGCGCGCGAGAGAGCGAGCTCCACCAAGCGAGCTCCCGCGCGCGCGCG 3000
QY 1001 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg 1020

Db 3001 CCACCGCAAAACCTGCAGCGCGAGAGAGAGCGCGCGCTCAGAGCTGTGGAGAGAGCGCGCG 3060
QY 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
Db 3061 GGCAGAGAGAGAGAGCGG 3120
QY 1041 LysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArg 1060
Db 3121 NNN 3180
QY 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro 1080
Db 3181 NNN 3240
QY 1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100
Db 3241 NGTCACCCAGCTGCGCGCGCTTCATGACATGCGCGCGCGCGCGCGCGCGCGCGCGCG 3300
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisAspSerValLeuGluArg 1120
Db 3301 ACCATCTCAACCGCGCTCCCTCATCTCTCTGTCCAAAGCACCGCGCGCTCTGAGAGG 3360
QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis 1140
Db 3361 CAATAGTGCATCTCCCAAGAGATGTGAGTCCAGCTCCACGTCGCCGTACTCAGAGCAT 3420
QY 1141 AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys 1160
Db 3421 GCCAAGGCGCGCGCGCGCGCTGTCAACATGAGGCTGCGCTGCGCATGAGACCCCAAAAG 3480
QY 1161 LeuAlaProPheSerGlyValLysGlnGluGluLeuSerProArgGlyGlnAlaGlyPro 1180
Db 3481 CTGGCACCTTTCAGCGAGATGAAGACAGAGACAGCTGTCCCAAGGAGCGAGCTGGGCGCA 3540
QY 1181 ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
Db 3541 CCGAGAGAGCTGGGGGGGCCACAGCGCGAGAGGCGTCCGTGTGAGAGAGGAGAGCTCTG 3600
QY 1201 GlySerValProGlyGlySerIleThrLysGlyTyrProSerThrArgValProSerAsp 1220
Db 3601 GGCTCAGTTCGGGCGGAGACATCACCAAGAGATTCGCCAGCACCGGAGTGCCTCGAGC 3660
QY 1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLys 1240
Db 3661 AGCGGCATCATACCGCGCTCATCACCGGACCGCGAGCTGTGTATACAG 3720
QY 1241 GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu 1260
Db 3721 GGCACCATCACAGATCATCGGCGAGAGACAGCCGAGTGTGTGACCGCGGCGGAG 3780
QY 1261 AspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyr 1280
Db 3781 GACAGCTGTGCCAAGGCGCGCATCTAGAAAGAGAGAGAGCGCATGCTGTCTAT 3840
QY 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro 1300
Db 3841 GAGGNN 3900
QY 1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArg 1320
Db 3901 NNN 3960
QY 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db 3961 NNN 4020
QY 1341 SerProHisIleLeuLysGluGlnHisIleValArgGlySerIleThrGlnGlyIlePro 1360
Db 4021 AGCGCCACCATCTCAAGAGAGAGACACATCTCGGCGGTCTTACACAGAGNNNNNNNN 4080
QY 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380

[illegible][illegible]

Qy	2101	AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu	2120
Db	6301	GGCCACCTCCACACCTCGGGCCGCTGCTTAGAGCCAGCCCTCGTCCAGCCCTGCTC	6360
Qy	2121	GlnThrAlaProGluValIleGlyHisGlnArgValValThrLeuAlaGlnHisGlnSer	2140
Db	6361	CAGACCGCCCAAGGGGTCAAGGTCCACAGCGGGTGTACCTCGGCCACACATCACT	6420
Qy	2141	GluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro	2160
Db	6421	NN	6480
Qy	2161	AlaProLeuTyrSerPheProGluAlaSerCysProValLeuAspLeuArgArgProPro	2180
Db	6481	NN	6540
Qy	2181	SerAspLeuTyrLeuProProProAspHisGlnAlaProAlaArgGlySerProHisSer	2200
Db	6541	NN	6600
Qy	2201	GluGlyGlyIleArgSerSerProGluProAsnIleYsThrSerValIleuGlyGlyIleAsp	2220
Db	6601	NN	6660
Qy	2221	GlyIleGlnProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla	2240
Db	6661	GGTATTGAACCTGTGTCCCAACCGAGGGCATGACGGAGCCAGGCACTCCGGAGTGTCT	6720
Qy	2241	ValTyrProLeuLeuTyrTyrArgAspGlyGlnGlnThrGluProSerArgMetGlySerIle	2260
Db	6721	GTTGACCCGCTGCTGTACCGGGAAATGGGGACAGCGGAGCCACAGATGGCTCCAG	6780
Qy	2261	SerProGlyAsnThrSerGlnProProAlaPhePheSerIleLeuThrGluSerAsnSer	2280
Db	6781	TCTCAGGCAACACACGACCGCCGACCTTCTTCAGCAAGCTGACCGAGCACTCC	6840
Qy	2281	AlaMetValIleYsSerIleYsGlnGlnIleAsnIleYsIleuAsnThrHisAsnArgAsn	2300
Db	6841	GCCATGTGTCAAGTCCAGAGACAGAGAAGTCAACAAGATGTGAACCCCAACCGGAT	6900
Qy	2301	GluProGluTyrTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr	2320
Db	6901	GAGCCTGAATTCANN	6960
Qy	2321	GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGlnHisGlnSerThrAsnMet	2340
Db	6961	NN	7020
Qy	2341	GlyLeuGlnAlaIleIleArgIleAlaLeuMetGlyIleYsTyrAspGlnTyrGluGluSer	2360
Db	7021	NN	7080
Qy	2361	ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla	2380
Db	7081	NN	7140
Qy	2381	MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly	2400
Db	7141	NN	7200
Qy	2401	GlyIleYsAlaIleValSerGlyArgProSerSerArgIleYsAlaIleYsSerProAlaProGly	2420
Db	7201	GGGAAGGCCAAGGTCTCTGCGAGACCCAGCGCCGAAGGCCAAGTCCCGGCCCGGGC	7260
Qy	2421	LeuAlaSerGlyAspArgProProSerValIleSerSerValHisSerGluGlyAspCysAsn	2440
Db	7261	CTGGCATCTGGGAGCCGGCCACCTCTGTCTCTCAGTGACTCGAGAGGAGACTCGAAC	7320
Qy	2441	ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAla	2457
Db	7321	CGCCGAGACGCGCTCACCAACCGGTGTGGAGAGACAGGCCCTGTCTCGCA	7371

AY12688		7013 bp	DNA	linear	GSS 16-DEC-2003
LOCUS					
DEFINITION	Mus musculus NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY12688				
VERSION	AY12688				
KEYWORDS	AY12688.1 GI:39768653				
SOURCE	GSS.				
ORGANISM	Mus musculus (house mouse)				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 7013) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J., Adams,M.D. and Carcilli,M. Interfering nonneutral evolution from human-chimp-mouse orthologous gene trios				
TITLE	Science 302 (5652), 1960-1963 (2003)				
JOURNAL	14671302				
PUBMED	2 (bases 1 to 7013)				
REFERENCE	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J., Adams,M.D. and Carcilli,M. Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
AUTHORS	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
COMMENT	Location/Qualifiers				
FEATURES	1..7013				
source	/organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" <1..>7013 /gene="NCOR2" /locus_tag="HCOM4640"				
gene					
ORIGIN					
Alignment Scores:					
Pred. No.:	0	Length:	7013		
Score:	7109.50	Matches:	1509		
Percent Similarity:	63.77%	Conservative:	59		
Best Local Similarity:	61.37%	Mismatches:	768		
Query Match:	53.80%	Indels:	125		
DB:	9	Gaps:	17		
US-09-522-753-5 (1-2517) x AY12688 (1-7013)					
OY	1 MeSeSerGlySerThrGlnLeuValAlaGlnThrTPraArgAlathrGluProArgTYRPro				
Db	1 ATGTCAAGATTCACACAGCCTGTGGCACAAACATGGGGGCTGCTGAGCCCGTACCACA				
OY	21 ProHisSerLeuSerTYRProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu				
Db	61 CCCCATGGCATCTCTTACCCGGTGCAATGCCCCGCCAACACAGACGTCGGGGCTCTT				
OY	41 GluTYRGlnHisHisSerArgAspTYRAlaSerHisLeuSerProGlySerIleIleGln				
Db	121 GAGTACCAACACCCCGGTGACTACACTCACCTGCACACCTGTCACCCGGTTCATCATCAG				
OY	61 ProGlnArgArgArgProSerSerLeuLeuSerGluPhedInProGlyAengluwargserGln				
Db	181 CCACAGAGGAGGGCGGCGCTCACTGCTGTGAGAGTTCCAGCTGGAGTGAACGNNNNNN				
OY	81 GluLeuHisLeuArgProGluSerHisSerTYRLeuProGluLeuGlyLysSerGluMet				
Db	241 MNN				
OY	101 GluPheIleGluSerLYSarGProArgLeuGluLeuLeuProAspProLeuLeuArgPro				
Db	300 NNN				

Oy		121	SerProLeuIleuAlaThrGlyGlnProAlaGlySerGluAspLeuTrpLysAspArgSer	140
Db		360	NNNAACCTGAGC	419
Oy		141	LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGluLeu	160
Db		420	CTGGCAGCGAAGCTGGAGCCGTGTGTCACTTCCAGTCCCCGCCACGCCTGACCCTTGAGCTA	479
Oy		161	GluLeuValProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAsp	180
Db		480	GAGCTGGGGCCATCTCGACTCTCCAGAAGGAGGAGTGATCCAGAAATGGAACGGGTGGAC	539
Oy		181	ArgGluIleThrMetValGluGlnGlnIleSerLysLeuIstlysLysGlnGlnGluLeu	200
Db		540	CGTGAGATCACCATGTGTGAGACAGACAGATCTCCAAGCTGAAGAAAGAACACAAACAGTTG	599
Oy		201	GluGlnGluValAlaLysProProGlnProGluLysProValSerProProIleGlu	220
Db		600	GAGAGAGGAGCGCGCAACCGCCGCAACCCAGAAAGCCTGTGTGCCACCAACCATTAAGAA	659
Oy		221	SerIsthisArgSerLeuValGlnIleIleIleTrpAspGluAsnArgLysLysAlaGluAla	240
Db		660	TCAAAGCACCGAAGCGTGTCTCAAGATATTCTACAGATGAGAACCGGNNNNNNNNNNNNNNN	719
Oy		241	AlaHisArgIleLeuGlnGlyLeuGlyProGlnValGluLeuProLeuTrpAsnGlnPro	260
Db		720	NNCTCTGTACAACAGCGC	779
Oy		261	SerAspTrpTrpArgGlnTrpHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu	280
Db		780	TCTGCACACAGCCAGTAGTACATGAAAAACATCAAAAATAAACACGAGCGATCGGAAGAAGCTG	839
Oy		281	IleLeuTrpPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln	300
Db		840	ATCTTGTACTTTAAAGCGAGAACCAACCGCGCAAGCATGTGGAAACAGCGCTTCTGCCAG	899
Oy		301	ArgTrpAspGlnLeuMetGluAlaLeuGlnLysLysValGluArgIleGluAsnAsnPro	320
Db		900	CGCATATGACCAAGCTCATGAGAGCGCGGGAAGAAGATAGAGCGCATAGAAACAATCCG	959
Oy		321	ArgArgArgAlaLysGlnSerLysValArgGluTrpTrpGlnLysGlnPheProGluIle	340
Db		960	CGAAGGAAGGGCCACAGAGAGACAGAGTGAGGAGTCTTCGAGAAACAAGTCCCAGAGATC	1019
Oy		341	ArgLysGlnArgGlnLeuGlnGlnLysrMetGlnSerArgValGlnArgGlySerGly	360
Db		1020	CGCAAGCAGCGGAGACTCAGAGAGCGCATGCAGACACAGAGGAGCGGCAACCGGCAAGTGG	1079
Oy		361	LeuSerMetSerAlaAlaArgSerGlnIstGlnValSerGluIleIleAspGlyLeuSer	380
Db		1080	CTCTCCATGTCGGCTGCCCGCAGATGAGATAGGTTTCTGAGATCATTTGATGCTGTGCT	1139
Oy		381	GluGlnGluAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPrometLeuTrp	400
Db		1140	GAGCAGGAGAACTGGAGAAAGCAGATGCGCCAGCTGCGCTGATCCCCCATGTTGTAC	1199
Oy		401	AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPromet	420
Db		1200	GACCGGGACACGACAGAGATCAAGTTCAACAACATGAATGACCTCATGATGACCCCAATG	1259
Oy		421	LysValIleTrpLysAspArgGlnValMetAsnMetTPSerGlnGlnGlnLysGlnLysPhe	440
Db		1260	AAGGCTTCAAGAGACCGTCAAGTTTCAACAATGTGGACCGACGAGAGAGGACACCTTC	1319
Oy		441	ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg	460
Db		1320	CGTGAGAAANN	1379
Oy		461	LysThrTrpAlaGluCysValLeuTrpTrpTrpLeuTrpLysAsnGluAsnTrpLys	480
Db		1380	NNNAAGGTCGCTGAATGTGTCTCTATTATTACTACCTGACCAAGAAAGAAATTAACAAG	1439

OY		481	SerLeuValAlaArgAsSerTyrTAArgATGAGTGLySereGlInGIInGIInGIInGIIn	500
Db		1440	AGCTTGCGAGCCGACACTATCGCGCGCGTAGGAAGCAGNNNNNNNNNNNNNNNNN	1499
OY		501	GInGIInGIInGIInGIInGIInGIInGIInProMecProArgSereSerGIInGIInLyS	520
Db		1500	NN	1559
OY		521	AspGIInLySGIInLySGIInLySGIInLysGIInGIInGIInLySProGIInValGIInAsn	540
Db		1560	NN	1619
OY		541	AspLySGIInAspLeuLeuLySGIInLyThAspAspThrSereGIInLyAspAsnAspGIIn	560
Db		1620	NN	1679
OY		561	LySGIInLysValAlaSereLySGIInArgLyThTrAlaAsnSereGIInLyAGAArgLySGIIn	580
Db		1660	AAAGAGGCCGTGGCTCCAAAGGCCGCAAACTGCACAAGCCAAAGCCCCCGCAAAGCC	1739
OY		581	ArgLIeThrArgSereMetAlaAsnGIInLysAlaAsnSereGIInLysAlaLIeThrProGIInGIIn	600
Db		1740	CGTATCACGGCGCTCCATGAGCCAAGAGGCCAACCATGAGAGACAGCCACCACAGCA	1799
OY		601	SerAlaGIInLeuAlaSerMetGIInLeuAsnGIInSereSerArgTrpThrGIInGIInGIInLysMet	620
Db		1800	AGTTCCAGAGCTGCGCTTCATGSAATGAAGAAGAGTTCTGCGTGAATTGAGGAAGAATG	1859
OY		621	GIInThrAlaLySlySGIInLeuLeuGIInHISGIInArgAsnTrpSereAlaLIeAlaArgMet	640
Db		1860	GAGACAGCAAAAGAAAGGCTCTCTGAAACATGGAGAGAACTGGTACGCCATTGCCCGCATG	1919
OY		641	ValGIInSeriLySThrValSeriGIInCyLeuAsnPheTyzrPheAsnTyzLySlysArgGIIn	660
Db		1920	GTGGGCTCCAGACCGGTGTCCAGTGAAGACTTACTTCAACTAACAAGAAAGAGCGAC	1979
OY		661	AsnLeuAspGIInLIeLeuGIInHIeLySleuLyMetGIInLySGIInLysAlaArgAlaArg	680
Db		1980	AACCTGACGAAATCTCTTCAGCACACAAAGCTTAAAGATGAGAGAGAGAGAAAGCGCTCGG	2039
OY		681	ArgLySlySlySlySlySlyAlaProAlaAlaSereGIInGIInLysAlaAlaPheProProVal	700
Db		2040	AGGAAGAAAGAAAGAACCCGACGTCGCGCAGCAGAGACAGGCTTCCACCTGCCGCT	2099
OY		701	GIInAspGIInLysMetGIInLysSeriGIInValSeriGIInAsnGIInGIInLysMetValGIInGIIn	720
Db		2100	CAGGACGAAAGATGGAAGCATCAGCGCGCAAGTCCCAATGAGGAAGAGCTGCGCGAGAG	2159
OY		721	AlaGIInLysAlaLeuHISAlaSereGIInAsnGIInValProArgGIInLyCSereGIInProAla	740
Db		2160	CGAGAGAGNN	2189
OY		741	ThrValAsnAsnSereAspThrGIInSeriLeProSerProHisThrGIInLysAlaAla	760
Db		2190	NN	2249
OY		761	AspThrGIInGIInAsnGIInProLySProProAlaThrLeuGIInLysAlaAspGIInProPro	780
Db		2250	GAG-----	2252
OY		781	GIInProProThrProProArgArgThrSereArgAlaProIIeGIInProThrProAlaSer	800
Db		2252	-----	2252
OY		801	GIInLysAlaThrGIInLysAlaProThrProProProAlaProProSerProSereAlaProPro	820
Db		2252	-----	2252
OY		821	ValValProLySGIInGIInLySGIInGIInLysThrAlaAlaAlaProProValGIInGIInGIIn	840
Db		2253	-----CAGAAgTCTGAGGCCGAGAG-----ATCGATGTGGCA	2285
OY		841	GIInGIInGIInLySProProAlaAlaGIInGIInLysAlaAlaAspThrGIInLySAlaGIInGIIn	860

Db 2286 AACCCAGAGAGCCCGAGGCTCTGAGAG-----CCCCCGAG 2324
Qy 861 ProValIysSerGlyThrGluGluGluGluProAla---LysGlyLysAsp 879
Db 2325 AGGTGAAGAGTGCACCAAGAGAGAGCCGAGAAAGCTTGAAGCAAAAGCCAAAGGCG 2384
Qy 880 AlaGluAlaIaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGly 899
Db 2385 ACAGAGGCCATTGAATGTGTCTGAGGCAACCTTAAGTGCAG-----GAGGCTGCT 2438
Qy 900 SerGlyArgAlaThrThr---AlaLysSerSerGlyAlaProGluAspSerSer 918
Db 2439 AGCAAGGCACTGTGACCAAGGGTTCCAGCTCAGTCCACCCAGACAGACATGACTCCAGT 2498
Qy 919 AlaThrCysSerAlaAspGluAlaAspGluAlaGluGlyLysLysLysAlaAspGluLeu 938
Db 2499 GCCACCTGAGTGCCTGATGAGTGAAGCAACCCGAGAGAGTGAAGAGGCGAGCTGCTG 2558
Qy 939 SerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAspAlaSerProGln 958
Db 2559 TCACCAAGGCCCAAGCTCTCCACCCGCGTGGAGATCCCGGCGCACTACCTGCCCCAG 2618
Qy 959 LysProLeuAspLeuLysGluLeuLysGluArgAlaAlaAlaIleProProIleGlnVal 978
Db 2619 AAGCCCTGAGCTGAAGCAGCTGAAGCAGCAGCAGCCCATCCCTATCCAGGTC 2678
Qy 979 ThrLysValAlaGluProProArgGluAspAlaAlaProThrLysProAlaProProAla 998
Db 2679 ACCAAGGTCATAGCCCTCCCGGAGAGACAGTACCCCAAGGATGTTCCCTGCTG 2738
Qy 999 ProProProGluAspLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer 1018
Db 2739 CCTCCACCCACGACGACCTACAGCCAGAGGTGACTGTCTCAGCAGTGGAGAGAGT 2798
Qy 1019 ProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaAlaGlu 1038
Db 2799 CCACGTGGCAAGTCCCGCAGCCAGCTGCTGCGAGAAAGAGNNNNNNNNNNNNNN 2858
Qy 1039 AlaGluLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValPro 1058
Db 2859 NNN 2918
Qy 1059 ProArgGluValIleLysAlaSerProAlaAlaProAspProSerAlaPheSerTyAla 1078
Db 2919 NNN 2978
Qy 1079 ProProGlyAlaIleProLeuProLeuGlyLeuAlaAspThrAlaArgProValLeuProArg 1098
Db 2979 NNN 3038
Qy 1099 ProProThrIleSerAspProProProLeuIleSerSerAlaLysAspProSerValLeu 1118
Db 3039 CCCCCC---ATCTTAACCCCCCAACCTCATCTCTGCAAGACATCCCGCGTACTT 3095
Qy 1119 GluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuAlaValProTySer 1138
Db 3096 GAGAGGCACTGGTGCATCTCCAGAGGATGTCAAGCTTGTGTGCTTCACTCA 3155
Qy 1139 GluAlaIleLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPro 1158
Db 3156 GAGCAAGCCAGGCCCCCATGGGCTCTCAACATGGGGCTGCCCTTGGCGGAGCCCT 3215
Qy 1159 LysLysLeuAlaAspPheSerGlyValLysGlnGluLeuSerProArgGlyGlnAla 1178
Db 3216 AAGAACTGAGCCCCCTTCAGTGAAGTGAAGCAAGAAAGTGTCTCCCTCGGGGTCAAGGCT 3275
Qy 1179 GlyProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuAspGlyThr 1198
Db 3276 GGGCCACCTGAAGTGTGGAGTGGCCACATGCTCAGAGAGACCTGTGTGCTGAAGGAGCA 3335
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218

Db 3336 GCACTGGGCTCCGCAACAGTGAAGCATCAAGAGGCTCCCACTACCCGGGCTGCA 3395
Qy 1219 SerAspSerAlaIleThrTyArgGlySerIleThrAlaGlyThrProAlaAspValLeu 1238
Db 3396 GAGGCCCC-----ACCTACAGAGCTTATACCCACGAGGAGCCCGCAGAGCTCTC 3449
Qy 1239 TyLysGlyThrIleThrArgIleIleGlyLysAspSerProSerArgLeuAspArgGly 1258
Db 3450 TACAGAGGTACATCAGCAGAGATGTCGTGAAGACAGCCCAAGTCCCTTGACCGGCA 3509
Qy 1259 ArgGluAspSerLeuProLysGlyLysValIleTyTrpGluGlyLysLysValLeu 1278
Db 3510 CGAAGGACACCTGCGCCCAAGGCGCATGTATATAGGAGCAAGAAAGCCACAGTCTTA 3569
Qy 1279 SerTyTrpGluGlyLysMetSerValThrGlnCysSerGlyLysAspGlyArgSerSerSer 1298
Db 3570 TCCTATGAAGNN 3629
Qy 1299 GlyProProIleGluThrAlaAlaProLysArgThrTyAspMetMetGluArgVal 1318
Db 3630 NNN 3689
Qy 1319 GlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu 1338
Db 3690 NNN 3746
Qy 1339 ArgHisSerProHisLysLeuLysGluGlnHisAlaIleArgGlySerIleThrGlnGly 1358
Db 3747 CAGCAGAGCCCC---CACTCAAGAGACACATCATCCGAGGCTCATCAGCAAGNN 3803
Qy 1359 IleProArgSerTyArgValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLeu 1378
Db 3804 NNN 3862
Qy 1379 LysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyLys 1398
Db 3863 NNN 3895
Qy 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
Db 3896 NNN 3955
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGlu 1438
Db 3956 NNN 4015
Qy 1439 LeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys 1458
Db 4016 NNN 4075
Qy 1459 TyrAspThrGlyAlaSerThrThrArgLysLysLysAlaArgSerLeuIleGly 1478
Db 4076 NNN 4135
Qy 1479 SerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAla 1498
Db 4136 AGCCCGGCGGCGCTTCCCTGCGCGACACCGCTGAGCAATATGTGTATGCCGGGCA 4195
Qy 1499 LeuGluArgAlaCysTyTrpGluGluSerLeuLysSerArgProGlyThrAlaSerSerSer 1518
Db 4196 CTGGAGCGTGCCTGATGAAGAGCTGGAAGAGCCGGTCAAGAGACACACATGTGTGCA 4255
Qy 1519 GlyLysSerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGln 1538
Db 4256 GGGGGCTCATTCACAGTGGGCTCCAGTGTGTGTGTGCTGAATGGGCAAGCCAGGCAA 4315
Qy 1539 SerProLeuThrTyArgLysAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySer 1558
Db 4316 AGCCCACTGACTTACGAAGACCAAGGAGCACTTCAACAGTCACTGCACTGCTCC 4375
Qy 1559 ProValThrMetCysArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLys 1578
Db 4376 CTTGTGACACAGAGGAGCCACGCGCTTCAAGAAAGNNNNNNNNNNNNNNNNNNNNNN 4435

QY 1579 AlaSerGlnAspArglyLeuThrSerThrProArgGluIleAlaIysSerProHisSer 1598
 DB 4436 NNN 4495
 QY 1599 ThrValProGluHisIleProHisProIleSerProTyGluHisIleuLeuArglyVal 1618
 DB 4496 NNN 4555
 QY 1619 SerGlyValAspLeuTyArgSerHisIleProLeuAlaPheAspProHisIlePro 1638
 DB 4556 NNN 4615
 QY 1639 ArgGlyIleProLeuAspAlaAlaAlaAlaTyTyTyLeuProArgHisIleuAlaProAsn 1658
 DB 4616 NNN 4675
 QY 1659 ProThrTyProHisIleuTyTyProProTyTyLeuIleArgGlyTyTyProAspThrAlaAla 1678
 DB 4676 CCACCTTACCCACACCTTACCCACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 4735
 QY 1679 LeuGluAsnArgGlnThrIleIleAsnAspTyTyIleThrSerGlnGlnMetHisIleAsn 1698
 DB 4736 CTGGAGAACCGGCAGACCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 4795
 QY 1699 ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
 DB 4796 GCTGCTCCGCCATGCGCCAGCGCTGTCATCTCAGGGGTCTGTCACTCCGAGAGTCTC 4855
 QY 1719 SerLeuAlaIleuAsnTyTyAlaAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro 1738
 DB 4856 TCGCTGGCCCTCAAT 4915
 QY 1739 HisLeuProValIleuValProProThrProGlyTyThrProAlaThrAlaMetAspArgLeu 1758
 DB 4916 CACCTGCCCGTGCTGTGTGCACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4975
 QY 1759 AlaTyLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer 1778
 DB 4976 GCCTACTCCCACTGCGCCCACTGCGCCCACTGCGCCCACTGCGCCCACTGCGCTCTC 5035
 QY 1779 ProGlyGlyProThrHisIleuThrTyProThrTyThrSerSerSerGlyuArgGlyuArg 1798
 DB 5036 CCAGGAGGCCCACTCACTAGCTAAACCAACCACTCACTCACTCACTCACTCACTCACT 5095
 QY 1799 AspArgAspArgGluArgAspArgAspArgGluArgGluSerIleuThrSerThr 1818
 DB 5096 GAACGTGACGGGAGAGAGAC-----AAGTCATCTCTACAGTCTACC 5137
 QY 1819 ThrThrValGluHisIleAspProIleTyArgProGlyTyThrGluGlnSerSerGlySerSer 1838
 DB 5138 ACTACAGCTGGAGATCACCCATCTGAGACCTGTAGCGAGCAGAGCAGCGGGCT---- 5194
 QY 1839 GlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisIleHis 1858
 DB 5195 -----GGGGGCGAGCGCGCGCGCGCTCCACAC-----CAC 5227
 QY 1859 GlnHisSerProIleSerProArgThrGlnAspAlaIleuGlnGlnArgProSerValIleu 1878
 DB 5228 CAGCACTGCGCCATCTCCCGGACCCAGAGCGCTTGACAGAGAGCGCCACATGTGCTG 5287
 QY 1879 HisAsnThrGlyMetIysGlyIleIleThrAlaValGluProSerIysProThrValIleu 1898
 DB 5288 CACAACACAGCATGAAGGCGGTGTCCTCTCGTGAACCGGCAACCGCCACGCTCTG 5347
 QY 1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
 DB 5348 AGGTTCACCTCCACCTTCTGCTGTCCGCCAGCTGCACATTCACCTCCACCTCCACCCAC 5407
 QY 1919 CysProLeuGlyGlyTyThrLeuAspGlyValTyTyProThrIleuMetGluProValIleu 1938
 DB 5408 TGCACCTTGATGTGACCCCTTGAAGGGGTCTACCTTACCTCATGAGAGCGGCTCGTGA 5467

QY 1939 ProTyGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
 DB 5468 CCCAAGAGACCTCTCGGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5527
 QY 1959 PheLeuAlaIysProProAlaArgSerGlyLeuGluProAlaSerSerProSerTyGly 1978
 DB 5528 TTCTTCACCAACCCCGGCGCG-----GAGCGCGCTCTCTACCCACGACAGAC 5578
 QY 1979 SerGluProArgProLeuValProProValSerGlyHisIleAlaThrIleAlaArgThrPro 1998
 DB 5579 TCCGAGCCCCGATCTCCCTAGACACCCCGAGCTCAGACGACAGCATGCGCGCACCCCA 5638
 QY 1999 AlaTyAsnLeuAlaProHisIleAlaSerProAspProProAlaProAlaSerAla 2018
 DB 5639 GCAAAGAGCTTGACACCCACCATGCTCCGAGCCCGCGCGCG---CCGACCTTGGCGC 5695
 QY 2019 SerAspProHisIleArgGlyuTyThrGlnSerTyProPheSerIleGlnGluLeuGluLeu 2038
 DB 5696 TCGAGTCTGCACCGAAGAAAGACTCAAGTAAACCTTTTCATCTCCAGAAATTGAACTC 5755
 QY 2039 ArgSerLeuGlyTyTyHisIleGlySerSerTySerProGluGlyValGluProValSerPro 2058
 DB 5756 CTTCTCTCGNN 5815
 QY 2059 ValSerSerProSerLeuThrHisAspTyGlyLeuProTyHisIleuGluGluLeuAsp 2078
 DB 5816 NNN 5875
 QY 2079 TySerHisIleuGluGlyGluLeuArgProTyGlnProGlyProValTyLeuGlyGly 2098
 DB 5876 NNN 5935
 QY 2099 GluAlaAlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSerSerPro 2118
 DB 5936 GAGCTGCCCATCTCCACATCTGCGGCACATGCGCGAGCAGCAGCAGCAGCAGCAGC 5995
 QY 2119 LeuLeuGlnThrAlaProGlyValIysGlyHisGlnArgValValThrIleuAlaGlnHis 2138
 DB 5996 CTCTCTCCAGACTGCCCGAGGACTCAAGAGGTCAACAGAGGTGTCACTGCTCGAC 6055
 QY 2139 IleSerGluValIleThrGlnAspTyTyThrArgHisIleProGlnGlnLeuSerAlaPro 2158
 DB 6056 ATCAGCANN 6115
 QY 2159 LeuProAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLeuArgArg 2178
 DB 6116 NNN 6175
 QY 2179 ProProSerAspLeuTyLeuProProProAspHisGlyAlaProAlaArgIysSerPro 2198
 DB 6176 NNN 6235
 QY 2199 HisSerGluGlyGlyTyAspArgSerProGluProAsnTyThrSerValIleuGlyGly 2218
 DB 6236 NNN 6295
 QY 2219 GluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArg 2238
 DB 6296 GAGATGCCATGAGCTGTGTCTCCACAGAGGGATGATCATGAGCAGACATGCTCG 6355
 QY 2239 SerAlaValTyTyProLeuLeuTyTyArgAspGlyGluGlnThrGluProSerArgMetGly 2258
 DB 6356 ACGCGTGTATCCCACTGCTGTATCGAGACGGGGAAACAGGGGAGCCCGAGGAGTGGG 6415
 QY 2259 SerTySerProGlyAsnThrSerGlnProProAlaPhePheSerTyLeuThrGluSer 2278
 DB 6416 TCTAAGTCTCCAGGCAACACAGCAGCGCCAGCTTTCAGTGAAGCTGAGTGAAGC 6475
 QY 2279 AsnSerAlaMetValIysSerTyTyGlnGluIleAsnTyLeuLeuAsnThrHisIleAsn 2298
 DB 6476 AACTCCGCAATGTGAAGTCAAGAAAGCAGAGATCAACAAGAACTCAACCCCAAC 6535
 QY 2299 ArgAsnGluProGluTyThrAsnIleSerGlnProGlyTyThrGluIlePheAsnMetProAla 2318

[illegible]

Pred. No.:	0	Length:	6836
Score:	6533.00	Matches:	1407
Percent Similarity:	61.81%	Conservative:	1
Best Local Similarity:	61.76%	Mismatches:	870
Query Match:	49.44%	Indels:	0
DB:	9	Gaps:	0

US-09-522-753-5 (1-2517) x AY412687 (1-6836)

QY	27	ProValGlnIleAlaArgThrIsthrAspValGlyLeuLeuGluIuTyrGlnIshIseSer	46
DB	2	CCAGTGCAGATTGCCCGGACGCACACGACGCTCGGGCTCTGGAGTACCAGACCACTCC	61
QY	47	ArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnAgaGArgPro	66
DB	62	CGCGACTAGCTCCACCTGCTCGCCCGGCTCATATCCAGGCCACGCGCGAGGCC	121
QY	67	SerLeuLeuSerGluPheGlnProGlyAenGluArgSerGlnIuLeuHisLeuArgPro	86
DB	122	TCCCTGCTGTCTGAGTTCACGCCCGGAATGAACGNNNNNNNNNNNNNNNNNNNNNN	181
QY	87	GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys	106
DB	182	NN	241
QY	107	ArgProAlaGluGluLeuLeuProAspProLeuLeuAlaArgProSerProLeuLeuAlaThr	126
DB	242	NN	301
QY	127	GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu	146
DB	302	NN	361
QY	147	ProValSerProProSerProBroHisThrAspProGluLeuGluLeuValProProArg	166
DB	362	CCGGGTCTCCCCCGACGCCCGGCACACCGCTGAGCTGGAGCTGGTCCGCCGCG	421
QY	167	LeuSerLysGluGluLeuIleGlnIshMetAspArgValAspArgGluIleThrMetVal	186
DB	422	CTGTCCAAGAGAGACTGATCCAGAAACAGACCGGTGACCGAAGATCACCAATGTA	481
QY	187	GluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaLys	206
DB	482	GAGCGACGAGTCTCTAAGCTGAAGAAAGACGANNNNNNNTGAGGAGGAGCTTCCAA	541
QY	207	ProProGluProGluLysProValSerProProProIleGluSerLysHisArgSerLeu	226
DB	542	CCGCCCGGACCTGAGAAAGCCCGTGTACCGCCGCCCATGAGTCCGAAGACCCGACCTG	601
QY	227	ValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaHisArgIleLeuGlu	246
DB	602	GTGCGAGATCATCTACGACGAGAAACCGNNNNNNNNNNNNNNNNNNNNNNNNNNNN	661
QY	247	GlyLeuGlyProGlnValGluLeuLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyr	266
DB	662	NN	721
QY	267	HisGluAsnIleLysIleAsnGlnAlaMetCArgLysLysLeuIleLeuTyrPheLysArg	286
DB	722	CATGGAACATCAAAATAAACACGAGCGATGCGGAAGAGCTAATCTTGTACTTCAAGAG	781
QY	287	ArgAsnHisAlaArgLysGlnThrLysGlnLysPheCysGlnArgTyrAspGlnLeuMet	306
DB	782	AGGAATCCGCTCGGAACCAATGGAGAGCAAGAGTTCTGCCAGCGCTATGACCAAGCTCAT	841
QY	307	GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu	326
DB	842	GAGGCTCGAGAGAGAGAGTGGAGCGATCCGAACAACCCCGGCGCGGCGCAAGAG	901
QY	327	SerLysValAlaArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu	346
DB	902	AGCAAGTGCAGGTACTTACGAGAAAGAGATTCCTGTGAGATCCGCAAGACGCGAGCTG	961

[illegible]

Db	2042	GCCTCGGGCGCTGACCGAAATGAGAGAGAGTGGTGGAGAGCGCTGAAGNNNNNNNNNNNN	2101
QY	727	SerGlyAsnGluValProArgGlyGlyCysSerGlyProAlaThrValAsnAsnSerSer	746
Db	2102	NN	2161
QY	747	AspThrGlySerIleProSerProHisThrCyluAlaAlaValAspThrGlyGlnAsnGly	766
Db	2162	NN	2221
QY	767	ProIysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProPro	786
Db	2222	NN	2281
QY	787	ArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaPro	806
Db	2282	NN	2341
QY	807	ThrProProProAlaProProProSerProSerAlaProProProValProIlyGlyGlu	826
Db	2342	NN	2401
QY	827	LysGluGluGluGluThrTrpAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPro	846
Db	2402	NAGAGAGAGAGAGACCGCAGCAGCGCCCCCGAGTGAAGAGAGAGAGAGAGAGAACCCCC	2461
QY	847	AlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlyCys	866
Db	2462	CGGNNNGAGAGCTGGCGAGTGGACACAGAGGAAGCCCGAGAGCCCGCTCAAGAGCGAGTGC	2521
QY	867	ThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThr	886
Db	2522	ACGGAGGAGGCCGAGAGAGGGCCCGCCAAAGGCAAGACGCCGAGGCCGCTGAGAGCTTACG	2581
QY	887	AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAla	906
Db	2582	GCCGAGGGGGGCGCTCAAGCAGAGAAGAAGAGGGCGGAGACCGCAGGGCCACACAGCC	2641
QY	907	LysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluVal	926
Db	2642	AAAGAGCTGGCGCGCCCCCAGCAGACGACCTCCAGTGCACCTCAGTGCAGACGAGAGGTG	2701
QY	927	AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThr	946
Db	2702	GATGAGGCGCAGGGCGCGGAGCAAAACCGAGCTGCTGCCCAAGGCCAGCGCTCTCACCC	2761
QY	947	ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGluLeu	966
Db	2762	CCGATGGCGACCCCCCGGCGCAATGCTCTCACCCCAAGGCCACGTGACCTTAACAGAGTGG	2821
QY	967	LysGlnArgAlaAlaAlaIleProProIleGlnAlaThrLysValHisGluProProArg	986
Db	2822	AAAGCAGCAGGGGCTGGCATCTCCCCCATCCAGGTCAACAAAGTCCATGAGCCCCCGCG	2881
QY	987	GluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeuGln	1006
Db	2882	GAGGAGCGAGCTCCCAACCAAGCAGCTCCCAAGCCCAACCGCCACCGCAAAACCTGACAG	2941
QY	1007	ProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPro	1026
Db	2942	CCGGAGACGACGCGCCCTCAGCAGCCTGGCGAGCGCCCCCGGGCGAAAGAGAGAGCCCG	3001
QY	1027	AlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro	1046
Db	3002	GCACCCCCGCCGACAG	3061
QY	1047	ProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSer	1066
Db	3062	NN	3121
QY	1067	ProHisAlaProAspProSerAlaPheSerTyralaProProGlyHisProLeuProLeu	1086

[illegible][illegible]

QY	1807	AspArgGluArgGluLysSerIleuLeuHisThrThrThrValGluHisAlaProIle	1826
Db	5342	NNNNNNNGAGCGGGAAAGAGCCATCTCCACGTCCACNNNNNGTGTGACAGCGCCCATC	5401
QY	1827	TPArgProGluYThrGluGlnSerSerGlySerSerGlySerGlyGlyGlyGly	1846
Db	5402	TGGAGNNNNNGTACAGAGAGAGCGGAGCAGCGGACAGCGGTCGAGCGGTGGGGG	5461
QY	1847	SerSerSerArgProAlaSerHisSerHisValHisGlnHisSerProIleSerProArg	1866
Db	5462	AGCAGCAGCGCGCGCGCTCCACATCCCANNNCCACAGACATCGCCATCTCCCTTNN	5521
QY	1867	ThrGlnAspAlaLeuGlnGlnIleArgProSerValLeuHisSerThrGlyMetLeuGlyLe	1886
Db	5522	NN	5581
QY	1887	IleThrAlaValGluProSerLeuProThrValLeuArgSerThrSerSerPro	1906
Db	5582	NN	5641
QY	1907	ValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAsp	1926
Db	5642	NN	5701
QY	1927	GlyValIleYrProThrLeuMetGluProValLeuLeuProGlyAlaProArgValAla	1946
Db	5702	NN	5761
QY	1947	ArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLeuProProAlaArg	1966
Db	5762	NN	5821
QY	1967	SerGlyLeuGluProAlaSerSerProSerIleGlySerGluProArgProLeuValPro	1986
Db	5822	NN	5881
QY	1987	ProValSerGlyHisValAlaThrIleAlaArgThrProAlaValAspLeuAlaProHisHis	2006
Db	5882	CCGGCTCTGGGCAAGCCACGCACTGCGCACCCCTGGAAAGAACTCGCACTCCACAC	5941
QY	2007	AlaSerProArgProProAlaProProAlaSerAlaSerAspProHisArgGlyLeuThr	2026
Db	5942	GCGAGCCCGGACCCCGCGCGCCACTGCTGAGCTCGAGCCCGGAGAAAGACT	6001
QY	2027	GlnSerIleYrProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTYrHisGlySer	2046
Db	6002	CAAGTAAGACCTTTTCCATCCAGGAAGTGAACCTCGGTCTCTGNNNNNNNNNNNNNN	6061
QY	2047	SerTYrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHis	2066
Db	6062	NN	6121
QY	2067	AspIleYrGlyLeuProIleHisLeuGluGluLeuAspIleYrSerHisLeuGluGlyLeu	2086
Db	6122	NN	6181
QY	2087	ArgProIleYrGlnProGlyProValIleYrLeuGlyGlyGlyAlaAlaHisLeuLeuProHis	2106
Db	6182	NN	6241
QY	2107	ArgProLeuProGluSerGlnIleProSerSerSerProLeuLeuGlnThrAlaProGlyVal	2126
Db	6242	CGGCGCGCTCGAGAGCGAGCGCTCGTCCAGCCGCGTCTCAACACCGCCCGAGGGTCC	6301
QY	2127	IleGlyGlyHisGlnArgValValThrIleuAlaGlnHisIleSerGluValIleThrGlnAsp	2146
Db	6302	AAAGGTCAACACCGGTGTGTCACTCGCGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNN	6361
QY	2147	TYrThrArgHisHisIleProGlnGlnLeuSerAlaProLeuProAlaProLeuTYrSerPhe	2166
Db	6362	NN	6421
QY	2167	ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTYrIleuPro	2186

[illegible]


```
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="IMAGE:5016291"  
/issue_type="lung, small cell carcinoma"  
/clone_id="NH_MGC_7"  
/lab_host="DH10B-R"  
/note="Vector: pOTB7"
```

ORIGIN

Alignment Scores:

```
Prod. No.: 4,36e-140 Length: 2000  
Score: 2566.00 Matches: 503  
Percent Similarity: 90.72% Conservation: 15  
Best Local Similarity: 88.09% Mismatches: 31  
Query Match: 19.42% Indels: 22  
DB: 3 Gaps: 1
```

US-09-522-753-5 (1-2517) x BC033087 (1-2000)

```
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20  
Db 273 ATGTGGGATTCACACAGCCCTGTGGACAGACGTGGAGGCCACTGAGCCCCGCTACCCG 332  
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40  
Db 333 CCCACAGCCTTCTCTACCCAGTGCAGATGCGCCGACGACACAGAGCTCGGCTCCTG 392  
Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60  
Db 393 GAGTACACAGACCACTCCCGGACTATGCTCCACCTGTGCGCGCTCATCATCAG 452  
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80  
Db 453 CCCAGGCGGAGGCGCTCTCTGCTGTCTGAGTTCAGCCCGGAAATGAACGATCCAG 512  
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100  
Db 513 GAGCTCCACCTGGCGGAGAGTCCCACTCACTGCGCGAGCGGGGAGTGAAGATGAGATG 572  
Qy 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProArgProLeuLeuArgPro 120  
Db 573 GAGTTCATTAAAGCAAGCCCGCTGCTGAGAGCTGTGCTGCTGACCCCTGCTGACCG 632  
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140  
Db 633 TCACCCCTGTGGCCACGGGCGAGCTGCGGATCTGAAGCTTCACCAAGACCTGAGC 692  
Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160  
Db 693 CTGACGGGCAAGCTGGAAACGGGTGTCCCCCGACCCCGGACACTGACCTGAGCTG 752  
Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180  
Db 753 GACTGTGTGCGGCGACCGCTGTCCAAAGAGAGCTGTATCCAGAACATGAGCCGTGTGAC 812  
Qy 181 ArgGlnIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200  
Db 813 CGAGAGATCACTGATGAGACAGACAGATCTTAAGCTGAGAAAGAAAGCAACACACTG 872  
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220  
Db 873 GAGAGAGAGGCTGCGAAGCCCGCGAGCTGAGAAAGCCCGTGTACCGCCCATGAG 932  
Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValAlaGluAla 240  
Db 933 TCAGAGACACCGCAGCCGTGTGAGATCATCTACACAGAAACCGGAAGAGCTGAGCT 992  
Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValAlaGluLeuProLysTyrAsnGlnPro 260  
Db 993 GCACATGGATTCTGGAAGGCTGGGGCCCAAGGTGAGAGCTGCGCTGTACACACAGCC 1052  
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
```

```
Db 1053 TCCGACACCCGCGAGTATCATGAGAACATCAATAAACCAGCGATCGGAGAACCTA 1112  
Qy 281 ILeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300  
Db 1113 ATCTTGACTTCAAGAGAGAGATACGCTCGGAAACAATGGAGCGAAGTCTGCGAG 1172  
Qy 301 ArgTyrAspGlnLeuMetGluAlaLeuGlnLysLysValGluArgIleGluAsnAspPro 320  
Db 1173 CGCTATACACAGCTCATGAGAGGCTGTGGAGAAAGAGTGGAGCCCATCGAAGAACACCCC 1232  
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340  
Db 1233 CGGCGCGCGGCCCAAGGAGACCAAGTGTGCGAGTACTACAGAAAGCAGTTCCCTGAGATC 1292  
Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360  
Db 1293 CGCAGAGCGCGAGCTGCGAGAGGCGATCAGACAGAGGAGGAGCGGAGCGGAGAGTGG 1352  
Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380  
Db 1353 CTGTCCATGTGCGCGCGCGCGAGAGACAGAGGTGTCAAGATCATCATGCTGCTCA 1412  
Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400  
Db 1413 GAGCAGAGAGAACTGGAGAAAGAGATGCGCAGCTGCGCTGATCCCGCCATCTGTAC 1472  
Qy 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420  
Db 1473 GAGCTGACACAGACCGCATCAAGTTCATCAACATGAACGCGCTTATGCGCGACCATG 1532  
Qy 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440  
Db 1533 AAGGTGACAAAGACCCCGAGTCAATGACATGTGAGTGAAGAGAGAAAGACCTTC 1592  
Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460  
Db 1593 CGGAGAAATTATGACATGCCAAGAACTTGTGCTGATCGCATCATCTGGAGAGG 1652  
Qy 461 LysThrValAlaGluCysValIleuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480  
Db 1653 AAGACAGTGTGCTAGAGCGCTCTTACTACTGACTGAAGAAATGAATGAATATTAAG 1712  
Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500  
Db 1713 AGCTGTGAGAGAGAGACTATCGGCCCGGCAAGAGCTGCTGCTGCTAAAGTTCCAG 1772  
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520  
Db 1773 CAAAAAAGATCACCGCGCGAGTAAAGAGCTCCAGCCGAGAAAGTTCTGCGCCAGAAAG 1832  
Qy 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValAlaGln 540  
Db 1833 CCACAGGCCAAGAAAGAGCGCTGCTCAAAAGCTCAGAAAGGTCAAAAGATCCACGCC 1892  
Qy 541 AspLysGlu-----AspLeuLeuLysGluLys 549  
Db 1893 AGAAAGCACTGCTCCAAAGGCACTGCGCAAGAAAGATAAGTGCATATAAAAGTA 1952  
Qy 544 -----AspLeuLeuLysGluLys 549  
Db 1953 ATAAAGTCTTTTGTGACTGTTAAAAAAA 1985
```

RESULT 5
BC020427 1362 bp mRNA linear HTC 03-JAN-2002
LOCUS Homo sapiens, clone IMAGE:4179307, mRNA.
DEFINITION BC020427
ACCESSION BC020427.1 GI:18043012
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (baaes 1 to 1362)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalobdc@bcm.tmc.edu
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,
Lawrence, S., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAK Plate: 12 Row: 1 Column: 23
This clone has the following problem: no polyA-tail.
Location/Qualifiers
1. 1362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4179307"
/tissue_type="Brain, anaplastic oligodendroglioma with
1p/19q loss"
/clone_lib="NCI CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 1,346-122 Length: 1362
Score: 2264.00 Matches: 431
Percent Similarity: 99.54% Conservative: 0
Best Local Similarity: 99.54% Mismatches: 2
Query Match: 17.13% Indels: 0
DB: 3 Gaps: 0

US-09-522-753-5 (1-2517) x BC020427 (1-1362)

QY 1676 ThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMet 1695
Db 63 ACGGCGCGCTGGAGAACCGGAGACCATCATATGACTACATCACTCTGGCAGCAGATG 122	
QY 1696 HisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro 1715	
Db 123 CACCAACAACGCGGCCACCGGCATGGCCAGCAGCTGATATGCTGAGGGGCTCTGCC 182	
QY 1716 ArgGlnSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSer 1735	
Db 183 CGGAGTCTCTGGTGGACTCACTACGCTGCGGGTCCCGGAGGCATCATCGACTGTCC 242	
QY 1736 GlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMet 1755	
Db 243 CAAGTGCACACACTGCTGTGCTGTGCTGCCCGCAGACACAGCACCAGCCGCGCATG 302	
QY 1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSer 1775	
Db 303 GACCGCTTGCTACTACCTCCCAACGCGGCCCAAGCCCTTTCAGACGCGCCACACAGCTCC 362	
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrTyrProThrThrThrSerSerSerGln 1795	
Db 363 CCACTCTCCCAAGAGGTCCAAACACTTGAACAAAACCAACCAACGACGTCTCTGTCGAG 422

QY 1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlyLeuSerIleLeu 1815
Db 423 CGGAGCGGAGACCGGGATCGAAGCGGGACCGGGATCGGGAGAGGAGAAAAGTCCATCTCC 482	
QY 1816 ThrSerThrThrThrValGluHisAlaProIleTTPArgProGlyThrGluGlnSerSer 1835	
Db 483 AGCTCCACCAACGAGGTGAGCAACGACCATCTGAGACCTGTATCAAGACAGAGACAC 542	
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855	
Db 543 GGCAGAGAGCGAGCAGCAGCGCGGGGGTGGGGGAGAGCAGCAGCGCCGCTCTCCACTCC 602	
QY 1856 HisAlaHisGlnHisSerProIleSerProAlaGthrGlnAspAlaLeuGlnGlnArgPro 1875	
Db 603 CATGCCACACACACTCGCCCATCTCCCTCCGAGCCACAGATGCTCTCAGACAGAGACC 662	
QY 1876 SerValLeuHisAsnThrGlyMetLeuGlyIleIleThrAlaValGluProSerLeuPro 1895	
Db 663 AGTGTCTTCACAAACACAGGATGAAAGGATATCATCCGCTGTGAGGCCACAGCAGCC 722	
QY 1896 ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProPro 1915	
Db 723 ACGTCTGAGGTCCACCTCCACCTCTCACCCTTCGCGCGGTGCGACATTCACACT 782	
QY 1916 AlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrIleMetGluPro 1935	
Db 783 GCACCCACTGCGCCACTGGCGCGGACCCCTCGATGGGGGTCTACCTCACTGAGGCC 842	
QY 1936 ValLeuLeuProValGluAlaProArgValAlaArgProGluArgProArgAlaAspThr 1955	
Db 843 GCTTGTGCTGCCAAGAGGAGGCCCCCGGGTCCCGGCGAGAGCGGCCCGGACAGACAC 902	
QY 1956 GlnHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPro 1975	
Db 903 GGCATAGCTTCTCTGCTGCCAAGGCCCGCCCTCGGGCTGAGGCCGCTCTCTCC 962	
QY 1976 SerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla 1995	
Db 963 ACGAAGGCTCGAGAGCCCGGCGCCCTAGTCTCTCTCTGCGCACCGCCACCATGCGC 1022	
QY 1996 ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro 2015	
Db 1023 CGACACCTCGCAAGAACTCTGCACCTCAACAGCCAGCCCGGAGCCGCGGCGCACCT 1082	
QY 2016 AlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGln 2035	
Db 1083 GCTTGGCTTCGGAACCGGACCGGGAAGAAAGACTCAAAAGTAAACCTTTCCATCCAGAA 1142	
QY 2036 LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro 2055	
Db 1143 CTGGAACCTCGTTCTCTGGGTTTACCAACGACAGCTACAGCCCGCAAGGGGTGAGGCC 1202	
QY 2056 ValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGln 2075	
Db 1203 GTCAAGCTGTGAGCTCACCTGACCTGACCAAGAGGGGCTCCCAAGCACTCGAA 1262	
QY 2076 GlnLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys 2095	
Db 1263 GAGCTGCAAGAAGCACTGAGAGGGGAGCTGCGGCCCAAGAGCAGAGCCCGGTGAAG 1322	
QY 2096 LeuGlyGlyAlaAlaHisLeuProHisLeuArgPro 2108	
Db 1323 CTGGCGGGAGGCGCCCACTCCACACTCGCGCGC 1361	

RESULT 6
BC026028
LOCUS BC026028 1966 bp mRNA linear HTC 22-MAR-2002
DEFINITION Homo sapiens, similar to nuclear receptor co-repressor 1, clone
IMAGE:3836580, mRNA.
ACCESSION BC026028
VERSION BC026028.1 GI:19664000
KEYWORDS HTC.

[illegible]

Sequencing Center (NISC) ,
Gatherersburg, Maryland:
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.ugc@nih.gov
Ahter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Boufield,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Khong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Mastello,C., Maekeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Statutidop,S., Thomas,P.J., Touchman,J.W.,
Turgeson,C., Vogt,J.L., Walker,M.A. Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L. H. and Green,E.D.

Clone distribution: MCC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Series: IRAK Project: 79 Row: M Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22538460
This clone has the following problem: retained intron.

Location/Qualifiers
1. 1850

```

FEATURES
  source
    Location/Qualifiers
      1..1850
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:5749816"
        /tissue_type="Pancreas, Spleen, adult pooled"
        /clone_id="NH_MGC_120"
        /lab_host="DH10B"
        /note="Vector: pCMV-Sport6"

```

Alignment Scores:			
Pred. No.:	1.97e-85	Length:	1850
Score:	1634.00	Matches:	346
Percent Similarity:	74.35%	Conservative:	57
Best Local Similarity:	63.84%	Mismatches:	96
Query Match:	12.36%	Indels:	43
DB:	3	Gaps:	12
US-09-522-753-5 (1-2517) x BC035748 (1-1850)			
QY	16	Gluprroarqtrgtyrproprohisserleuserfyrprovalgluilealaqthrhithr	35
Db	272	CAAGTGGTATGCTCTCCACCTGTCAGATATCATTTCCACACCCGCCACACAG	331
QY	36	Aspvalglyleuengluyrgln-----Hisiserarqasryrlaserrhisleu	53
Db	332	GAGTTCGACGTCCTCGATTATGCTCTCTCATTTTAAGTAGCTACGGCATCAGCTT	391
QY	54	Serprogliserllelelglnproglinarqarqproserleuleusergluphegl	73
Db	392	TTTGAGCAACAGCAGCAGCAACAGCTTCGAAGGGACCTTCCTTGCTTTCAGATTTTCA	451
QY	74	Progllyasngluarqserglngluen--His-LeuayrProgluserhis-----Se	90
Db	452	CCAGGTCTGACAGAGTGTATTTTCTTTGATGATATTAGGCTTCACAGAAAGCAACATAG	511
QY	90	rlyrleuPro---gluengllyssserglumetgluphe-----llegluserlysar	107
Db	512	TTTATGAACCGTTTCATCCAGGCCCATCCCACTGATCATGATTCACCTGGATCGAAGCG	571
QY	107	qProarqleugluuleuProasp-----	116
Db	572	ACACGCTCGGAACAGGTTTCTGATTCATTTTCACGCTGCAGTGTGCGGTTTGGC	631
QY	116	OleuLeuayrProserProleuLeualaThrlyglInProAlaglysergluapleuth	136
Db	632	TTTATGTCACCCGCTGCCA-----GAAGGGCTGAGGACT--TCTGCAGATGCTAA	679
QY	136	rlysarqarqserleuthrhlylylsleuglu---ProvalserProproserProprohi	155
Db	680	GAAGGATTCACGATTCGAGGCAAACTGAAGCTCATCTCTTCAATTTTCGGGGCAAC	739

[illegible]

```

Db      1769  ACCCTCCGACGAGAAAAAGTACGAGAAAAAGAAAGAGGATTAACGACGAAAAACAGAAAA 1828
Oy      534   uLyE 535
Db      1829  AAAA 1832

RESULT 8
BC029627
LOCUS   BC029627
DEFINITION Mus musculus, similar to nuclear receptor co-repressor 1, clone
IMAGE:4501757, MRNA.
ACCESSION BC029627
VERSION   BC029627.1
KEYWORDS  GI:20988194
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Scturognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS  Strauberg, R.
TITLE     Direct Submmission
JOURNAL  Submitted (06-MAY-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: gcgabs-remail.nih.gov
          Tissue Procurement: The Cepko Laboratory
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
          DNA Sequencing by: Baylor College of Medicine Human Genome
          Sequencing Center
          Center code: BCM-HGSC
          Web site: http://www.hgsc.bcm.tmc.edu/cdna/
          Contact: amg@bcm.tmc.edu
          Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
          Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
          Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov
Series: IRAX Plate: 31 Row: K Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers
1. 1835
   /organism="Mus musculus"
   /mol_type="mRNA"
   /db_xref="taxon:10090"
   /clone="IMAGE:4501757"
   /tissue_type="Eye, retina, mouse strain C57BL/6"
   /clone_id="N1H MGC_94"
   /lab host="DH10B"
   /note="Vector: pCMV-Sport6"

FEATURES
source
1. 1835
   /organism="Mus musculus"
   /mol_type="mRNA"
   /db_xref="taxon:10090"
   /clone="IMAGE:4501757"
   /tissue_type="Eye, retina, mouse strain C57BL/6"
   /clone_id="N1H MGC_94"
   /lab host="DH10B"
   /note="Vector: pCMV-Sport6"

Alignment Scores:
Pred. No.: 5,37e-85 Length: 1835
Score: 1626.50 Matches: 342
Percent Similarity: 73.43% Conservative: 56
Best Local Similarity: 63.10% Mismatch: 93
Query Match: 12.31% Indels: 51
DB: 3 Gaps: 12

US-09-522-753-5 (1-2517) x BC029627 (1-1835)
Oy      16  GUBPARGTGTTPROPCHISERLSEUSETTYPRVAGLNLHLEALAGTTHRIETHR 35
          :::::
          282 CAAAGTCGTATATCTTACATCTGTGTCACGATATACCTTTCCAGTACCCGACACCGACG 341

```


AUTHORS Strauberg, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe and Dr. Mina Cerna Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 BC Cancer Agency, Vancouver, BC, Canada
info@bcgc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Teal, Natesja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 131 Row: c Column: 2
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES
 source

1..1819
 /location/Qualifiers
 1..1819
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30240594"
 /tissue="Jaw bone, mouse, day 10.5 to 11.5 (CD-1), developing maxilla and mandibula tissue containing undifferentiated progenitor cells for muscle, dermis, epidermis, skin, membranous bone, cartilage and teeth"
 /clone_lib="NIH_MGC_164"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"

ORIGIN

Alignment Scores:

Pred. No.: 6..94e-85 **Length:** 1819
Score: 1624.50 **Matches:** 341
Percent Similarity: 73.94% **Conservative:** 59
Best Local Similarity: 63.03% **Mismatches:** 91
Query Match: 12.29% **Indels:** 50
DB: 3 **Gaps:** 11

US-09-522-753-5 (1-2517) x BC062162 (1-1819)

QY 16 GlnProArgProProHisSerLeuSerTyProValGlnIleAlaArgThrHisThr 35
 Db 282 CAAAGTCGATCTTCACATCTGTCACATATACCTTCCACGATACCGACACGACG 341
 QY 36 AspValGlyLeuLeuGluTyrgln-----HisHisSerArgAspTyralaSerHisLeu 53
 Db 342 GAATTTGCAAGTTCCTGACACTCGCTTCATATGAAATTGACGACGATACAGCTT 401
 QY 54 SerProGlySerIleIleGlnProGln--ArgArgArgProSerLeuLeuSerGluPhe 72
 Db 402 TTACACAGCAGCAGCAGCAGCAGCAGCTTCGAAAGCAGACCTTCCTTGCTTCAGAATT 461
 QY 73 GlnProGlyAsnGlnArgSerGlnGluLeu-HisLeuArgProGluSerHisSerTyrlle 92
 Db 462 CACCCGGGTTCTGACAGCTCTCAAGAAAGAAAGAACTGATGAACATTTTCACTCA 516
 QY 92 uProGluLeuGlyLysSerGluMetGluPhe-----IleGluSerLysArgProArgle 110

Db 517 -----GACCCCTCACCGGATCATGATTCCTTGAGTCGAAGGCCACGCCCT 566
 QY 110 uGluLeuLeuProAsp-----ProLeuLeuAr 119
 Db 567 GAGACGAGTTTCTGATGCGCATTTCCAGCGTGTAGTTCGGGTTTACCTTATGATCA 626
 QY 119 gProSerProLeuLeuAlaThrGlyGlnProAlaGly-----SerGluAspLeuTh 136
 Db 627 C-----TCGCGCCAGAAAGCTTGAAGCTTGAAGCTTCGACAGTGTCTAA 665
 QY 136 rLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProSerProProHis 155
 Db 666 GAAGSATTACCATTTTGAAGCAAAACATGAAGCTCCATCTCTCTTGGCTGGGCAACC 725
 QY 155 sThrAspProGluLeuGluLeuValProProAlaGlyLeuSerGlyGluLeuIleGlnAs 175
 Db 726 ATGTGAGATGACCAAAATGCTTCAAGCTTCAAGCTTCAAGAGAGATTATACAGAG 785
 QY 175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195
 Db 786 TATGACCGGGTGAAGCCGAGAGATTTGCAAAAGTAGACAGACATCTTAAATTGAAAA 845
 QY 195 sLysGlnGlnLeuGluGluGluLysAlaLysProProGluProGluLysProValSe 215
 Db 846 GAAACAAACACAGCTAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
 QY 215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyArgProGluAnar 235
 Db 906 CCTCTCTCTGCGTGAACAAACACCGTATATGTCGCAAAATTAATATATATGAGAAATCG 965
 QY 235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGluGlyLeuGlyProGlnValGluLeuPr 255
 Db 966 GAAAAAGCAGAAAGAGCTCATAAATATTGTAAGCTCTTGCCCAAAATTTGAACCTGCC 1025
 QY 255 oLeuTyArgGlnProSerAspThrArgGlnTyrglnHisGluAsnIleLysIleAsnGlnAl 275
 Db 1026 ACTCTAATACACAGCATCAGATGACCAAGGTGATCAGTACAGACATCAAGCAAAACCCAGT 1085
 QY 275 sMetArgLysLysLeuIleLeuTyThrLeuValGluGluGluValAlaArgLysGlnTrpLys 295
 Db 1086 GATGAGAGAAAACTCATTTTATTTTAAATATATATATATATATATATATATATATATAT 1145
 QY 295 sGlnLysPheCysGluArgTyrglnArgGlnLeuMetGluAlaLeuGluLysLysValGluAr 315
 Db 1146 ACAAATAATCTGCACGCTATGATCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
 QY 315 gIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrglnTyrglnTy 335
 Db 1206 AATGAGAAATTAATCCCGGAGGAAAGCAAAAGAAACCAAAAGAGAAATATATGAGAA 1265
 QY 335 sGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlu 355
 Db 1266 GCAGTTTCCAGAAATTCGAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1322
 QY 355 yGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIle 375
 Db 1323 TCAGAGGGAGAGCTGCTTTCACGCCACATGCTTGAAGATGACATGAGATTTCTGAAT 1382
 QY 375 eIleAspGlyLeuSerGluGlnGluLeuAsnLeuGluLysGlnMetArgGlnLeuAlaValIle 395
 Db 1383 TATTGATGCTCTTTCGAAACAGAGAAATATGAGAGAAATATGCGTCAAGCTTCTGAT 1442
 QY 395 eProProMetLeuTyArgAlaAspGlnGlnAlaGlyIleLysPheIleAsnMetAsnGlyLe 415
 Db 1443 TCACCTATATGATTTGATGACGAAACAAAGAGGCTCAATTCATATATATATATATATAT 1502
 QY 415 nMetAlaAspProMetLysValTyrlsAspArgGlnValMetAsnMetTrpSerGluGlu 435
 Db 1503 GATGAGAGATCCATTAAGCTTTATTAAGACAGACGTTTATGAACTTTGACATGACCA 1562
 QY 435 nGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAl 455

Db 1563 TGAAGAAGATCTTTAAGACAGATTATTCAGCATCCAAAACTTTGACTAATTGC 1622
QY 455 aSerPheLeuGluAlaGlySerThrValAlaGluCysValLeuTyrTyrTyrLeuThrIleGly 475
Db 1623 ATCCATTATTTGAAGAGAGTGCTCTGATTTGTTTATTTACTATTATTAACCAAGAA 1682
QY 475 sAaenGluAntYrLysSerLeuValAlaArgSerTyr---ArgArgArgGlyLysSerG 494
Db 1683 AATATGGAATTTAAAGCCCTCCTGAGAGAAATATGAAAAACGACAGACGAATAATCA 1742
QY 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 1743 GCAGCAG-----ATTGCCCG 1757
QY 514 gSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGlu 534
Db 1758 TCCCTCACAAGAAAGAAAGTGAAGAAAGAAAGAAAGTAAAGCAAGAAAAA 1817
QY 534 u 534
Db 1818 A 1818
RESULT 10
LOCUS BM909096 1093 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6611948 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431719
5', mRNA Sequence.
ACCESSION BM909096.1 GI:19359475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1093)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgarbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1904 row: 9 column: 16
High quality sequence stop: 741.
Location/Qualifiers
1..1093
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5431719"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_98"
/note="Organ: Brain; Vector: pOTB7, Site1: XhoI, Site2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
Pred. No.: 3 58e-85 length: 1093
Score: 1624.00 Matches: 331
Percent Similarity: 92.78% Conservative: 3
Best Local Similarity: 91.94% Mismatches: 14

Query Match: 12.29% Indels: 12
DB: 5 Gaps: 3
US-09-522-753-5 (1-2517) x BM909096 (1-1093)
QY 2151 HisPProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer 2170
Db 19 CACCACAGCAGCTCAGGCACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78
QY 2171 CysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHis 2190
Db 79 TCCCCCGTCTGAGACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 138
QY 2191 GlAlaProAlaArgLysSerProHisSerGlnGlyLysAspSerProGluProAsn 2210
Db 139 GGTGCCCCG 198
QY 2211 LysThrSerValLeuGlyGlyGlyGluAspGlyLysGluProValSerProProGlnGly 2230
Db 199 AAGACGTGGTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 258
QY 2231 MetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlu 2250
Db 259 ATGACGAGCCAGGCGACTCCCGAGTGTGTACCGCTGTACCGGATGGGAA 318
QY 2251 GlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAla 2270
Db 319 CAGACGAGCCAGCAGCAGATGGGCTCCAAGTCTCCAGGCAACACACAGCCGCGCAGCC 378
QY 2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLeuGlnGly 2290
Db 379 TTCTTCAGCAACCTTACCGAGAGCACTCCGCAATGCTCAATCCAGAAAGAGAGATC 438
QY 2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGly 2310
Db 439 AACAGAAAGCTAACACCCCAACCGAATGAGCTGAATACATATACGCCAGCTGGG 498
QY 2311 ThrGluLeuPheAsnMetProAlaIleThrGlyLysMetThrTyrArgSerGln 2330
Db 499 ACGAGATCTTCATATGATGCGCGCGCATCACCGGACGGCTTATGACCATATGAAGCCAG 558
QY 2331 AlaValGlnGlnHisAlaSerThrAsnMetGlyLysGluAlaIleLeuArgLysAlaLeu 2350
Db 559 GCGGTGCGAGAACATGCGCACCAACATGAGGCTGAGGCGCATATTTGAAGAGCATTC 618
QY 2351 MetGlyLysTyrAspGlnThrGluGluSerProProLeuSerAlaAsnAlaPheAsnPro 2370
Db 619 ATGGGTAAATATGACCAAGTGGGAAGATCCCGCGCTCAGCGCCATCTTTAAACCT 678
QY 2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSer 2390
Db 679 CTGAATGCCAGTGCACAGCTGCGCGCTGATGCCATTAACCGCTGACCGACGAGAT 738
QY 2391 AspHisThrLeuThrSerProGlyGlyGlyGlyValAlaLysValSerGlyArgProSer 2410
Db 739 GACCAACACACTCACTCCGCAAGTGGCGCGGGAAGCAAGAGTCTCTGGAGAGCCACAC 798
QY 2411 SerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVal 2430
Db 799 AACCGAAAGCCAAAGTCCCGGAGCCCGGAGTGGCATCTGGGGAACCGGCACCTCTGTC 858
QY 2431 SerSerValHis-SerGlnGlyAspCys-AsnArgArgThrProLeu-ThrAsnArgVal 2449
Db 859 TCTCTAGTCACTTGGAGAGAGATCGAACAACGCGGAGCGCGCTCAACAACGCGCTG 918
QY 2450 -TrpGlu-AspArgProSerSerAlaGly-SerThrPro-PheProTyrAsnProLeu 2468
Db 919 GTGGGAAGACAGCGCCCTTCTCCGAGGTTTCCCGCATTTTCCCTTAACCCCGCAT 978
QY 2468 eMetArgLeuGlnAlaGlyValMet-----AlaSerProProProGlyLeuProAl 2486
Db 979 CARGCGGTTGACAGGGGGGAGTTCATGGGTTTCCCAACCCCAACGCGGCGCTCCCGCG 1038

QY 2486 agly---SerGlyProLeuAla---GlyProHis-HisAlaATP 2498
DB 1039 CGGGGAAAGCGGGCCCTTCGCGTGGCCCCCACCACCGCTGG 1082

RESULT 11
LOCUS BC026623
DEFINITION Mus musculus, similar to nuclear receptor co-repressor 1, clone IMAGE:4981672, mRNA.
ACCESSION BC026623
VERSION BC026623.1 GI:20073007
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1996)
Serausberg, R.
Direct Submission
Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey B. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) medpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRM Plate: 59 Row: m Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: frame shifted.
Location/Qualifiers
1. 1996
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4981672"
/rname="Colom, normal. 5 month old male mouse."
/clone_id="NCI CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 1.45e-83 Length: 1996
Score: 1603.00 Matches: 337
Percent Similarity: 73.40% Conservative: 52
Best Local Similarity: 63.58% Mismatches: 91
Query Match: 12.13% Indels: 50
DB: 3 Gaps: 12

US-09-522-753-5 (1-2517) x BC026623 (1-1996)
QY 16 GIUProArgTyProHisSerLeuSerTyProValGlnIleAlaArgThrHisThr 35
DB 490 CAAAGTCGTTATCTTCACATTCGTGCATATACCTTCCAGTACCGACACGAG 549
QY 36 AspValGlyLeuLeuGlnTyGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
DB 550 GAATTTCAGTTCCTGATACCGCTCTTCATATTGAAGTTAGCCAGCATCAGACTT 609

QY 54 SerProGlySerIleIleGlnProGln---ArgArgArgProSerLeuLeuSerGluPhe 72
DB 610 TTACAGCAGACAGACAGACAGACAGACAGCTTCGAAAGACACTTCCTGTTCAAGATTT 669
QY 73 GlnProGlyValGlnTyGlnSerGlnIleLeuHisLeuAlaArgProGluSer---HisSerTy 91
DB 670 CACCCGGGTTTCAGACAGCTTCAGAAAGAGAAAGAAATGATATGAAACGTTTCACTCA--- 726
QY 92 LeuProGluLeuGlnTyGlnSerGluMetGluPhe-----IleGluSerTyValArgProArg 109
DB 727 -----GAGCCCTCAGCGGTGATCATGATTCCTTGAGTCCAGCCGCCAGC 774
QY 110 LeuGluLeuLeuProAsp-----ProLeuLeu 118
DB 775 CTGAGCCAGTTCCTATGATGCCCTTCCAGCGGTGATGCTCCGCTTTTACCTTTAGTT 834
QY 119 ArgProSerProLeuLeuAlaThrGlnProAlaGly-----SerGluAspLeu 135
DB 835 CAC-----TCGCTGCAGAAAGGCTTGAGTGTGACAGATGCT 873
QY 136 ThrTyAspArgSerLeuThrGlyTyGlnGlu---ProValSerProProSerProPro 154
DB 874 AAGAGGATTCACCATTTGGAGACCAACATGAAGCTCCATCTCTCTTGGCTGGGCA 933
QY 155 HisThrAspProGluLeuGluValProProArgLeuSerTyGlnIleGln 174
DB 934 CCATGTGAGATGACCAAAATGCTTCACTTCAAAAGCTTTCAGAGAGATTATACAG 993
QY 175 AsnMetAspArgValAspArgGluIleThrMetValGlnGlnIleSerTyGlnVal 194
DB 994 AGTATGAGCCGGGTACCGAGAGATTGCAAAAGTAGAGAGACAGATCTCTTAATTGAAA 1053
QY 195 TyGlnGlnGlnIleGlnIleGlnIleGlnIleAlaTyProProGluTyProVal 214
DB 1054 AAGAAACAAACACGCTAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
QY 215 SerProProProIleGlnSerTyHisArgSerLeuValGlnIleIleTyAspGluAsn 234
DB 1114 TCCCTCTCTCTGTGAAACAAACACCGTATGATGTCCTCAAAATTTTATGATGAGAA 1173
QY 235 ArgTyGlnAlaGlnIleAlaHisArgIleLeuGlnTyGlnIleProGlnValGluLeu 254
DB 1174 CGGAAAGAGACAG 1233
QY 255 ProLeuTyArgGlnProSerAspThrArgGlnTyHisGluAsnIleLeuValGln 274
DB 1234 CCACTCTATACAGACATCAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1293
QY 275 AlaMetArgTyGlnIleLeuTyPheTyArgValArgAsnHisAlaArgTyGlnIle 294
DB 1294 GTGATAGAGAAAGATCTCATTTTATTTTAAAGAGAAATCATACAGAAACAAAGG 1353
QY 295 TyGlnTyPheCysGlnArgTyArgGlnLeuMetGluAlaLeuGlnTyGlnValGlu 314
DB 1354 GAAACAAAGAAATCGCCAGCGCTATGATGATGATGATGATGATGATGATGATGATGATG 1413
QY 315 ArgIleGlnAsnAsnProArgArgAlaTyGlnSerTyValArgGlnTyTyGln 334
DB 1414 AGAATGAGAAATATATCCCGAGAGAAAGCAAAAGAAAGCAAAAGCAAAAGCAAAAGCA 1473
QY 335 TyGlnPheProGlnIleArgTyGlnArgGlnLeuGlnIleGlnTyGlnIleArgVal 354
DB 1474 AAGCAGTTTCAGAAATTCAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1530
QY 355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlnIleGlnValSerGln 374
DB 1531 GGTCAAGAGGAGAGGTGCTTTCAGCCACATGCTGTAAGAGAGATGATGATGATGATGAT 1590
QY 375 IleIleAspGlyLeuSerGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 394
DB 1591 ATTATTGATGCTCTTTCAG 1650

[illegible]

FEATURES

```

/organism="Homo sapiens"
/mol_type="tRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5470428"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH1 MGCC 41"
/notes="Organ: skin; Vector: pOTB1; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAC(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NH1 MGCC Library."

```

ORIGIN

Alignment Scores:	
Pred. No.:	1.95e-81
Score:	1550.00
Percent Similarity:	95.00%
Best Local Similarity:	95.00%
Query Match:	11.80%
DB:	4
	length: 965
	Matches: 30
	Conservative: 0
	Mismatches: 11
	Indels: 3
	Gaps: 3

US-09-522-753-5 (1-2517) x BM555371 (1-965)

QY	2182	AspLeuTyrLeuProProProProAsnHisGlyValaProAlaArgGlySerProHisSerGlu	2201
Db	3	GACCTTACCTCCGCCGCCCGGAGCAGTgTgCCCCGGGCGGTgGCTCCCGCCACAGCGAA	62
QY	2202	GlyGlyIysAspSerProGluProProAsnIleThrSerValLeuGlyGlyGlyIysAspGly	2221
Db	63	GGGGGCAAGAGGTCTCCAGAGCCAAACAGACGTCGGTCTTGGGTGGTGGAGGACGGT	122
QY	2222	IleGluProValSerProProGluGlyMetThrGluProGlyHisSerThrSerAlaVal	2241
Db	123	ATTGAACCTGTGTCCCCACCGAGGGCAGTGCAGAGCCAGGGCACTCCCGGAGTGTGTG	182
QY	2242	TyrProLeuLeuTyrATGAspGlyGlyGluInThrGluProSerATGMetGlySerIysSer	2261
Db	183	TACCGCTGCTGTACCGGATGGGAGACAGACGAGCCAGACAGATGGCTCCAAgTCT	242
QY	2262	ProGlyAsnThrSerGlnProProAlaPhePheSerIysLeuThrGlySerAsnSerAla	2281
Db	243	CCAGGCAACACCAAGCCAGCCCGCAGCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCC	302
QY	2282	MetValIysSerIysIysGlnGluIleAsnIysIysLeuAsnThrHisAsnAspAsnGlu	2301
Db	303	ATGTGCAAGTCCAGAGACCAAGATCAACAAGAGCTGAACACCCACACACCGAGAAATAG	362
QY	2302	ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly	2321
Db	363	CCTGAATTCATATACAGCCAGCCGAGGAGCGAGATCTTCATATATGCGCGCATACCGGA	422
QY	2322	ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly	2341
Db	423	ACAGGCTTATAGACTATATAGAAAGCCAGCGGTGAGGAACTGTGCAGACCAACATGGGG	482
QY	2342	LeuGluAlaIleIleArgIysAlaLeuMetGlyIysTyrAspGlnTrpGluGlySerPro	2361
Db	483	CTGGAGCCATATATTGAAGAGCACTACGGGTAAATATGACCACTGGGAAGAGTCCCCG	542
QY	2362	ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMet	2381
Db	543	CCGCTCAGCGCCAATGCTTTTAAACCTCTGAATGCCAGTGCACAGCTGCCGCTGTATG	602
QY	2382	ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly	2401
Db	603	CCCAATAACCGCTGCTGAGCGAGCGAGTATCAACACACTCAGCTGGCCAGGTGGCGGCGG	662
QY	2402	LysAlaIysValSerGlyArgProSerSerArgIysAlaIysSerProAlaProGlyLeu	2421
Db	663	AAGGCCAAGTCTCTGCGAGACCCAGCACCGGAAAAGCCAAGTCCCGGCCCGGGCTGTG	722
QY	2422	AlaSerGlyAspArgProProSerValHisSerGluGlyAspCysAsnArg	2441
Db	723	GCATGTGGGGACCGGCCACCTCTGTCTCTCCAGTGCACCTGCGAGGGGAGACTGCACACGC	782
QY	2442	ArgThrProLeuThrAsnArgValITrGluAspArgProSerSerAlaGlySerThrPro	2461
Db	783	CGAGAGCGGCTCACCAACCGGGGTGGAGAGACAGGCCCTGTGTCGACGGGTCCAGCGCA	842
QY	2462	PheProTyrAsnProLeuIleMet-ArgLeuGlnAlaGly--ValMetAlaSerProPr	2480
Db	843	TTCCCTTCAAAACCCCTGATCATGGCCTGACAGGCGGGAGTGCATGGTATTTCACCAAC	902
QY	2480	oProProGly--LeuProAlaGly-----SerGlyProLeuAlaGlyProHis 2496	

Db 903 CCCAGCGGGGCTTCCCGCGGGGAGCGGGGCCCCCTCGGCTGGGCGCCCGCAC 960

RESULT 13
BMS53310 1126 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT_6572642 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467163
DEFINITION 5', mRNA sequence.
ACCESSION BMS53310
VERSION BMS53310.1 GI:18791930
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS 1 (bases 1 to 1126)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cga@rs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LICM1968 row: 1 column: 12
High quality sequence stop: 674.
Location/Qualifiers
1. 1126
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5467163"
/rlnsue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 41"
/note="Organ: skin, Vector: pOTB7, Site 1: XhoI, Site 2: EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	3,18e-80	Length:	1126
Score:	1540.00	Matches:	314
Percent Similarity:	82.1%	Conservative:	9
Best Local Similarity:	79.90%	Mismatches:	16
Query Match:	11.65%	Indels:	55
DB:	4	Gaps:	1

US-09-522-753-5 (1-2517) x BMS53310 (1-1126)

QY 1457 LeuLysrThrApThrGlyAlaSerThrThGlySerIleuYshIaApValaIatSerLeu 1476

Db 2 CTCAGTACGACACCGCGCGCTCCACACCTGGCTCCAAAAGCAGACGACGCTCCCTC 61

QY 1477 ILGGLSerProGlyArgThrPheProProValaHisProIleuSerValMetAlaApAla 1496

Db 62 ATGGGAGCCCGCGCGCGGAGCTTCCACCCGTGACCCGCTGATGATGGCCGAGCGCC 121

QY 1497 ArgAlaLeuGluArgAlaCyTYrGluGluSerIleuSerIleuSerArgProGlyThrAlaSer 1516

Db 122 CGGGACATGGAACGTGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181

QY 1517 SerSerGlyIleSerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro 1536

Db 182 AGCTCGGGGGGCTCATTTGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241

QY 1537 ArgGlnSerProLeuThrTYrGluAspHisGlyValaProPheAlaGlyHisLeuProArg 1556

Db 242 CGGAG 301

QY 1557 GlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer 1576

Db 302 GGTTCGCGCGTGAACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361

QY 1577 SerIleValaSerGlnAspArgIleuThrSerThrProArgGlyIleAlaLysSerPro 1596

Db 362 AGCAGGATCCAG 421

QY 1597 HisSerThrValProGluHisIleProHisProIleSerProTYrGluHisLeuLeuArg 1616

Db 422 CACAGCAGCGTCCGAG 481

QY 1617 GlyValSerGlyValaAspLeuTYrArgSerHisIleProLeuValaPheAspProThrSer 1636

Db 482 GCGGTGAGTGGCGTGAACGTGATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541

QY 1637 IleProArgGlyIleProLeuAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1656

Db 542 ATACCCCGCGCATCCCTCTGAG 601

QY 1657 ProAspProThrTYrProHisLeuTYrProProTYrLeuIleArgGlyTYrProAspThr 1676

Db 602 CCCAGCCCGCATCCCGAG 661

QY 1677 AlaAlaLeuGluValaAspArgGlnThrIleIleAspAspTYrIleThrSerGlnIleMetHis 1696

Db 662 GCGGCGGTGAG 721

QY 1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716

Db 722 CACAGAGCGGCGAG 781

QY 1717 GluSerSerLeuAlaLeuAsnTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGln 1736

Db 782 GAGTCTCGCTGGCATCTCACTACGCTGCGGAGTCCCGCA----- 820

QY 1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp 1756

Db 820 ----- 820

QY 1757 ArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArgHisSerSerPro 1776

Db 820 ----- 820

QY 1777 LeuSerProGlyProThrHisLeuThrIleProThrIleThrIleSerSerSerGluArg 1796

Db 821 -----GAGAGTCCGAG 870

QY 1797 GluArgAspArg-AspArgGluArgAsp-ArgAspArgGluArgGluArgSerIleLeuT 1816

Db 871 GAGCAG 930

QY 1816 HisSerThrThrVal-GluHisAlaProIleTYr-ArgProGlyThrGluGlnSerSe 1835

Db 931 CGTCCAGCAG 990

QY 1835 rGlySerSerGlySerSerGlyIleGlyIle 1845

Db 991 CGGAGGATCCGAG 1021

RESULT 14

BM477568 1057 bp mRNA linear EST 05-FEB-2002

LOCUS AGENCOURT_6484946 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554420

DEFINITION 5', mRNA sequence.

ACCESSION BM477568

VERSION BM477568.1 GI:18526610

KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1057)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph. D.
Email: cga@db-rc@mail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12272 row: p column: 05
High quality sequence stop: 674.
Location/Qualifiers
1..1057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5554420"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC 85"
/note="Organ: Lymph; Vector: pCMV-SPORT6; Site 1: NciI; Site 2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC library."
ORIGIN
Alignment Scores:
Pred. No.: 9,02e-77 Length: 1057
Score: 1480.50 Matches: 308
Percent Similarity: 87.57% Conservative: 2
Best Local Similarity: 87.01% Mismatches: 27
Query Match: 11.20% Indels: 20
DB: 4 Gaps: 6
US-09-522-753-5 (1-2517) x BM477568 (1-1057)
QY 929 AAlAGlUGlYgLYAsPlySAsnArGLeUsErProArGProSeRleUeUThrProTr 948
DB 1 GCCGAGGGGGGGGCAAGAAACCGGCTGCTGTCCCAAGGCCCACTCTCCACCCGACT 60
QY 949 GLYAsPProArGLAAsnAAsErProGlnLYsProLeUAsPLeUlySGlnLeUlySGln 968
DB 61 GCGGACCCCGGGGCAATGCTCACCACCAAGCACTGGAAGCAGTGAAGCAG 120
QY 969 AArgAlAAlAAlAAlEProProlIeGlnValIThrLYsValHISgluProProlAArg 988
DB 121 CGAGCGCGCTGCAATCCCCCATCTCAGGTACCAAAAGTCCATAGAGCCCCCGAGAGAC 180
QY 989 AAlAAlAProThrLYsProAlAProProlAProProlProGlnInsLeUglnProGln 1008
DB 181 GAGGCTCCCAAGCAAGCAGCTCCCAAGCCCAAGCCGCAAAACCTGCGAGCCGGAG 240
QY 1009 SeRAsPAlAProGlnGlnProGlySeRSeRProArGLYLYsSeRAlYsSeRProAlAPro 1028
DB 241 AGGAGAGCCCTCAGACAGCTGCGACAGCCCGGGGCAAGAGCAGAGCCGCGACCC 300
QY 1029 ProAlAAsPlySGlu-----AlAheAlAAlAgluaIaGln 1040
DB 301 CCGCGGACAAAGAGGCAAGAAAGCTGTGTCTTCTCCAGCTTCCAGCCGAGGCCAG 360
QY 1041 LYsLeUProGLYAsPProProCYsTrPThSeRGLYLeUProPheProValProProArG 1060
DB 361 AAGCTGCTGGGAGACCCCTTCTGTGACTTCCGGCTGCTTCCCGGCCCCCGCT 420
FEATURES
source
Location/Qualifiers
1..1057

QY 1061 GlUValIILeLYAAlAsErProHlAAlAPrAsPProSeRAlAphSeRTrAlAProPro 1080
DB 421 GAGGTGATCAAGACCCCTCCCGCATGCGCGAGACCCCTCGCTTCTCCATCCCTCACCCT 480
QY 1081 GLYHsPProLeUProLeUgLYeUHLsAsPTrAlAArgProValLeUProArGProPro 1100
DB 481 GGTCAACCACTGCGCTGGGCTCCATGACATGACCTGCCGGCCGCTCTGCCGCCACCC 540
QY 1101 ThrILSeRAsPProProProLeUHLSeRSeRAlYsHISProSeRValLeUglnArg 1120
DB 541 ACCATCTCCAAACCCGCTCCCTCATCTCTCTGCAAGACACCCAGCTCTCGAGAGG 600
QY 1121 GlnILleGlyAlAAlEsErGlnLYsSeRValGlnLeUHLsValProTYrSeRGLuHIS 1140
DB 601 CAATAGGTCATCTCCCAAGAAATGTGTGTCAGCTTCACAGTCCCTGATCAGAGCAT 660
QY 1141 AAlYsAlAProValGLYProValThRMeTGLYsLeUProLeUProMeArSPProLYsLYs 1160
DB 661 GCGAAGGCCCCGGTGGGCTGTACATAGGGGCTGCCCTGCCATGAGACCCCAAAAG 720
QY 1161 LeUAlAProPheSeRGLYValLYsGlnGlnLeUSeRProArGLYglnAglYPro 1180
DB 721 CTGGACACCTTCAGCGAGTGAAGCAGAGCAGCTGTCCCAAGGGCCAGGCTGGGCCA 780
QY 1181 ProGlnSeRLeUgLYValProThrAlAglnglnAAlAsErValLeUArgGLYThrAlALeU 1200
DB 781 CCGGAGAG-CTGGGGGTGCACA-GCCAGAGAGCGTCTGTGAGAGGACAGCTCTTG 838
QY 1201 GLYSeRValProGLYgLYSeRILeThrLYsGLYILeProSeR---ThrArgValProSeR 1219
DB 839 GGCTCAAGTTCGGCGGGAGCATCACCAAGGCAATTCACAGACAGCGGGGCTCGG 898
QY 1220 AsPSeR-ALAlEThrTYrArg-GLYSeRILeThrHISGLY---ThrProAlAAsPValL 1238
DB 899 GACAGGNCATCATCATACCGCGGTTCCATCACCAAGGNAAGCCAGCTGAGCTGC 958
QY 1238 eUrTYrLYsGLYThr---ILeThrArgILe---ILeGLYglnAsPSeRProSeRArgLeU 1256
DB 959 TGGG-CMAAGCNCACCATCACAGGATCATCGGGCGAAGACAGNCCGAGTGCCTTG 1017
QY 1256 sPArg-----GLYArgGLYAsPSeRLeUProLYs 1265
DB 1018 GAAACNCCGCGGAGGAACAGCCCTGCGCCCAAA 1055
RESULT 15
BM423558 1051 bp mRNA linear EST 29-JAN-2002
LOCUS AGENCOURT 6399411 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5516745
DEFINITION 5', mRNA sequence.
ACCESSION BM423558
VERSION BM423558.1 GI:18391770
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1051)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph. D.
Email: cga@db-rc@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM2017 row: n column: 10
High quality sequence stop: 667.
FEATURES
source
Location/Qualifiers
1..1051

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5516745"
/libsource_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(C). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NH_MGC library."
```

ORIGIN

Alignment Scores:

```
Pred. No.: 1.02e-76 Length: 1051
Score: 1479.50 Matches: 305
Percent Similarity: 78.95% Conservative: 10
Best Local Similarity: 76.44% Mismatches: 24
Query Match: 11.20% Indels: 60
DB: Gaps: 4
```

US-09-522-753-5 (1-2517) x BM423558 (1-1051)

```
QY 1457 LeuLeuYrYrAhpThrAlaSerThrThrGlySerIlybHisAspValAArgSerLe 1476
Db 2 CTCAGAGTACGACTCCGGCGGCTCCACACCTGGGCTCCAAAAAGACACGATACCTCCCT 61
QY 1476 uileGlySerProGlyYrGThrPheProProValHisProLeuAspValMetAlaAspAl 1496
Db 62 CATCGGAGCCCGGCGGAGAGTTCCACCCGCTGACCCGCTGATGTAGCGCGAGCC 121
QY 1496 aATgAlaLeuGluAArgAlaCyTYrGluGluSerLeuIySerArpProGlyYThralase 1516
Db 122 CCGGGACACTGGAAGCTCCTGCTACGAGAGAGCTGAGAGCGCGGACAGGACCGCAG 181
QY 1516 rSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLyPr 1536
Db 182 CACTCGGGGGGCTCCATTGCGCGCGGCGCGCCCGCTGATGTGCTAGCGGGCAAGCC 241
QY 1536 oATrGlnSerProLeuThrTYrGluAAspHisGlyAlaProPheAlaGlyHisLeuProAr 1556
Db 242 GCGGACAGAGCCCTTAACCTATGAGAGCACAGGGGACACCTTTGCCGGCACCTCCACG 301
QY 1556 gGlySerProValThrMetAArgGluProThrProAArgLeuGlnGlySerLeuSerSe 1576
Db 302 AGGTTGCGCCGTGACACGCGGAGGCCACGCGCGCTGAGAGGAGGCAAGCTTTCGTC 361
QY 1576 rSerIyAlaSerGlnAspAArgLybLeuThrSerThrProAArgIuIleAlaIySerPr 1596
Db 362 CAGCAAGGACATCCAGAGACGAAAGCTGACGCTGACGCTGTGAGATCCCAAGTCCCC 421
QY 1596 oHisSerThrValProGluHisAspProHisAspProIleSerProTYrGluHisLeuLeuAr 1616
Db 422 GCAACAGACCTGCGCGAGACACCAACCCCATCTGCGCTTATAGACACTGCTTCG 481
QY 1616 gGlyValSerGlyValAspLeuTYrAArgSerHisIleProLeuAlaPheAspProThrSe 1636
Db 482 GGGCGTGAAGTGGGTGACCTGTATGACAGCACATCCCGCTGACCTTCGACCCCACTC 541
QY 1636 rIleProAArgLyIleProLeuAspAlaAlaAlaIleTYrTYrLeuProAArgHisLeuAl 1656
Db 542 CATATCCCGGGGATCCCTGTGACGAGCGCGCTGCTACTACCTGCCGACACCTGGC 601
QY 1656 aProAspProThrTYrProHisLeuTYrProProTYrLeuIleAArgGlyTYrProAspTh 1676
Db 602 CCCCAACCCCACTACCTGACCTGTACCACTTACTCATCCGCGCTACCCCGACAC 661
QY 1676 rAlaAlaLeuGluAspAArgGlnThrIleIleAspTYrIleThrSerGlnGlnMetHis 1696
Db 1676 rAlaAlaLeuGluAspAArgGlnThrIleIleAspTYrIleThrSerGlnGlnMetHis 1696
```

```
Db 662 GCGCGCGCTGAGAACCGGACGACCATCATCATATGACTACCTCGACGAGATGCA 721
QY 1696 rHisAsnThrAlaThralaMetAlaGlnAArgAlaAspMetLeuAArgGlyLeuSerProAr 1716
Db 722 CCACAAAGCGGACCCGCAATGCGCCAGCGAGCTGATATGCTGAAGGGGCTTCGCCCG 781
QY 1716 gGlySerSerLeuAlaLeuAsnTYrAlaAlaGlyProAArgGlyIleIleAspLeuSerG 1736
Db 782 CGAGTCTCGGTGCACTCACTACGCTGCGGGTCCC----- 818
QY 1736 nValProHisLeuProValLeuValProProThrProGlyThrProAlaThralaMetAs 1756
Db 818 ----- 818
QY 1756 pArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArghIsSerSerPr 1776
Db 818 ----- 818
QY 1776 oLeuSerPro-GlyGlyProThrHis-LeuThrLybProThrThrThrSerSerSerGlu 1795
Db 819 -----CCAGAGAGGCAACCACTTACCAACCAACCAACCAACCAACCAACCAACCA 871
QY 1796 AArgGluAArgAArg--AspAArgGluAArgAArgAArgGlu-AArgGlyLySerI 1814
Db 872 CCGGAGCGGAGAACCGGAGATCCAAAGCGCGGACCGGAGATCGGGAACCGGAAAGTCCAT 931
QY 1814 eleuThr---SerThrThrThralaGluHisAlaProIleTrp-ArgProGlyYThrGlnG 1833
Db 932 CTTTCCTTCCCCCAACGAGGGGAAAGCAAGCCCATCTGGAACCTGTGTAACCAAC 991
QY 1833 InSerSerGlySerSerGlySer-----SerGlyGlyGly 1845
Db 992 CCAACAACGGGACACAGGSCACTTACACTTCGCGGGGGGGGGG 1038
```

Search completed: November 12, 2005, 03:26:38
Job time : 24452 secs

THIS PAGE BLANK (USPTO)